RUNNING HEAD: USER PREFERENCES AND APPRAISAL PURPOSE

User Preferences For Appraisal Systems With Different Purposes

John R. Hundley

Indiana University South Bend

FINAL THESIS ACCEPTANCE RECORD SHEET
User Preferences For Appraisal Systems With Different Purposes
John R. Hundley, B.S.B.A., M.B.A.

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Master of Arts in Applied Psychology.

Frank Fujita, Ph.D.

Thesis Committee

Richard L. Gottwald, Ph.D.

Steven D. Norton, Ph.D.

Date of Oral Examination

April 26, 2001

Table of Contents

Abstract

A client organization found no consensus to adopt a proposed performance appraisal system because prospective users differed whether the system purpose should be developmental or administrative. The organization decided to survey user preferences. User reactions are the least researched appraisal system criteria, and preferences are the least studied reaction. No previous studies have measured user preferences for appraisal systems with different purposes.

Appraisal purpose shapes appraisal system design. Administrative systems employ between-person comparisons using global, norm-referenced rankings. Developmental systems employ within-person comparisons using dimensional, criterion-referenced ratings. Some experts argue the two purposes are incompatible, while others believe they can't be separated.

The researcher developed an appraisal system preference scale using 187 students. The scale was unidimensional with good reliability. Students rated four appraisal systems: administrative only, developmental only, combined, and separated. Combined systems pursue both purposes simultaneously while separated systems pursue both purposes sequentially. In study 1, students preferred developmental only or combined systems and least preferred administrative only systems.

In Study 2, the researcher surveyed prospective client organization users, presenting the same four appraisal systems. Of 274 users, 78 responded. Users rated, ranked, and chose alternatives from paired-comparisons. Users preferred developmental only or combined systems and least preferred administrative only systems.

User preferences explain the lack of consensus for adopting a proposed separated system. The researcher recommends a developmentally-focused, combined system to the client organization. The low response rate for Study 2 limits its generalizability. Future research could study reasons for appraisal system preferences.

User Preferences For Appraisal Systems With Different Purposes

The client organization for the current study is a regional campus of a multi-campus, state university system. The state university system decentralizes the design of staff compensation systems and related performance appraisal systems. The client organization does not have a merit pay system for staff. Staff pay increases granted at the start of each fiscal year have been based on an across-the-board percentage plus increases based on length of service. The client organization does not have a performance appraisal system for staff, however, employees with faculty appointments are appraised by their supervisors for merit pay increases. Because the client organization wished to explore the possibility of basing annual staff increases at least partially on merit, a Performance Evaluation and Merit Pay Task Force was formed. The Task Force membership included representatives from the biweekly and professional Staff Councils as well as the Academic Administrative Council. Mohrman, Resnick-West, and Lawler (1989) recommend such an approach as it builds commitment to a new system.

The Task Force conducted two surveys of the staff who would be rated and the administrators who would rate staff performance if a merit system was adopted. The first survey was a performance appraisal design questionnaire and the second survey was a merit pay system design questionnaire. The response rates to the two surveys were relatively low: 34.5% for the performance appraisal design questionnaire and 28.6% for the merit system design questionnaire.

The performance appraisal design questionnaire asked respondents to choose 4 out of 13 possible purposes for a performance appraisal system. The 13 potential purposes

were grouped into three categories as follows:

(1) strategic (linking employee effort to campus goals, linking employee effort to departmental objectives, and communicating campus values to staff);

(2) developmental (clarifying individual goals and expectations, identifying individual training needs, providing developmental feedback, coaching performance improvement, and determining individual career progression goals); and

(3) administrative (linking pay to performance, making promotion decisions, making retention decisions, and evaluating human resource systems).

The top ranked purposes included providing developmental feedback (61.8% of respondents selecting) and linking pay to performance (59.6% of respondents selecting). When asked which of the appraisal purpose categories were most important, 49.4% selected developmental, 39.3% selected administrative, and 11.2% selected strategic. This pattern of responses was similar for both supervisors and non-supervisors and by functional classification (i.e., biweekly staff, professional staff, academic administrator). These responses indicated support for a system with both administrative and developmental purposes.

In response to a statement that performance ratings should be used to determine merit increases, 84.3% responded either "strongly agree" or agree". In response to a statement that the evaluation form should include an overall rating that could be used to link pay to performance, 75.3% responded either "strongly agree" or agree". In response to a statement that the evaluation form should discuss the ratee's strengths and weaknesses, 88.8% responded either "strongly agree" or agree". In response to a

statement that the evaluation form should describe employee development needs, 82.0% responded either "strongly agree" or agree". This pattern of responses was similar for both supervisors and non-supervisors and for all functional classifications. These responses also indicated support for a system with both administrative and developmental purposes. As recommended by Mohrman et al. (1989), the survey responses were used to develop comprehensive specifications for a performance evaluation and merit pay system. The system specifications were developed to reflect the preferences for specific system design features selected by a majority of survey respondents. Based on the survey responses, the Task Force developed specifications for a "separated" performance appraisal system with parallel administrative and developmental purposes. The specifications called for separate administrative and developmental appraisal discussions to be conducted several months apart.

Because the proposed specifications represented a major organizational change, they were presented for in-depth discussion at meetings of the constituent groups affected. The system specifications were presented to the support and professional Staff Councils, the Academic Administrative Council, the Chancellor's Staff, and several divisional meetings. None of these meetings resulted in a consensus to adopt the proposed performance evaluation and merit pay system specifications. A major concern expressed in these meetings was that the proposed system was too burdensome. There were questions whether the benefits of the proposed system warranted the effort required for its implementation. Some meeting participants suggested eliminating the developmental aspects of the proposed system. Other meeting participants suggested focusing on a

developmental purpose and not linking pay to performance.

To resolve this dilemma, the client organization decided to conduct a preference poll presenting potential raters and ratees with a choice of appraisal systems with different purposes. The choices offered included systems with: (1) a developmental purpose only; (2) an administrative purpose only; (3) combined developmental and administrative purposes conducted at the same time; and (4) separated developmental and administrative purposes conducted in parallel at different times (i.e., the Task Force proposal). The two previous surveys asked respondents to choose among detailed appraisal system design features which had not been combined into integrated whole systems. The goal of the preference poll was to focus the choice on whole systems rather than components.

The challenge presented by the client organization was thus to measure user preferences for appraisal systems with different purposes. To meet this challenge, the performance appraisal literature was reviewed with two different foci: (1) assessing user preference for appraisal systems; and (2) designing appraisal systems with different purposes.

Assessing User Preference For Appraisal Systems

Murphy and Cleveland (1995) refer to user reaction criteria as the "neglected criteria" in appraisal research (p 310). Although most research has focused on the psychometric characteristics of the performance measures, employee opinions about the appraisal process may be as crucial or more crucial to its long-term effectiveness (Dipboye & de Pontbriand, 1981; Keeping & Levy, 2000; Waldman, 1997). Bernardin and Beatty (1984) advocated that future research should emphasize appraisal satisfaction due to its

critical role in determining appraisal effectiveness. Thomas and Bretz (1994) stated user acceptance research is virtually nonexistent even though user acceptance is a major concern of practicing managers in America's largest companies. Reaction criteria place a ceiling on the effectiveness of appraisal systems - user acceptance is a necessary, but not sufficient condition for an appraisal system to be effective (Murphy & Cleveland, 1995; Hedge & Borman, 1995).

Measuring User Reactions

There are many ways to operationalize employee reactions to performance appraisals (Cawley et al, 1998; Keeping & Levy, 2000). Roberts (1992) asserted that one of the serious weaknesses in performance appraisal research is the lack of a clear and consistent definition and operationalization of appraisal system acceptance. Several different types of dependent variables have been used to assess reactions to performance appraisals including perceived utility, fairness, accuracy, performance improvement, and satisfaction (Giles & Mossholder, 1990; Keeping & Levy, 2000). Satisfaction appears to be the most frequently used reaction measure (e.g., Dorfman, Stephan, & Loveland, 1986; Mount, 1984; Russell & Goode, 1988; Silverman & Wexley, 1984). Giles and Mossholder (1990) noted that one of the advantages of satisfaction measures is that they assess both fairness cognitions and simple affect (i.e., feelings and emotions). This affords a broader indicator of individuals' reactions (Lam & Schaubroeck, 1999) than more specific, cognitively oriented criteria such as perceived utility (Giles & Mossholder, 1990).

In their meta analysis of 27 field studies investigating the relationship between

participation and reactions to performance appraisal, Cawley et al. (1998) classified and coded reaction measures as one of the following categories: satisfaction, motivation to improve, utility, fairness, and "other". Measures coded as satisfaction were further divided into satisfaction with the appraisal session (interview) and satisfaction with the appraisal system, when possible. This is consistent with previous research suggesting that some aspects of an appraisal may be differentially related to these two different types of satisfaction (e.g., Giles & Mossholder, 1990; Mount, 1984). Although the current study is concerned with user preferences for appraisal systems, a review of the appraisal system satisfaction literature is warranted because user preferences should be associated with user appraisal system satisfaction.

There is no widely agreed upon scale by which fairness perceptions should be measured and there are also no standard questionnaires for the measurement of appraisal satisfaction (Cardy & Dobbins, 1994). A number of studies have used single-item measures of appraisal satisfaction (e.g., Bannister, 1986; Jordan, 1990; Landy, Barnes, & Murphy, 1978). The classic study by Landy et al. (1978) included both fairness and accuracy in the same single-item measure. Hedge and Teachout (2000) commented that a single-item measure of user acceptance is too simplistic to represent a broad, multifaceted construct. Roberts (1992) concluded that user acceptance consists of more than a single attitude, but rather represents a series of distinct components that combine to create a global attitude or affective orientation. Patten (1998) observed that most attitudes are complex constructs consisting of many elements. She advised using multiple items in an attitude scale. She also advised including only one point per item (e.g., fairness or

accuracy, but not both). The current study seeks to develop a multiple item appraisal system preference scale.

Thomas and Bretz (1994) concluded that real organizational concerns, such as how different appraisal system designs affect employee attitudes, have not been properly studied. Satisfaction studies, such as those reference above, ask for reactions to an appraisal system that has been implemented rather than a system being proposed.

Assessing User Preferences

Waldman (1997) remarked that a possible limitation of reactions research is that reactions represent an attempt to gauge beliefs and attitudes after the fact. Reactions seek to measure perceptions or degree of satisfaction with an existing system. An assessment of user reactions to an appraisal system only addresses the status quo and is inherently limited to what that system is offering (Waldman, 1997). In contrast, preferences allow for the study of broader design issues and possibilities (Waldman, 1997). Acceptance of an appraisal system is largely dependent on how the system incorporates user preferences and expectations (Manshor & Kamalanabhan, 2000). Bretz, Milkovich, and Read (1992) criticized the fact that appraisal systems are typically designed by personnel specialists with little or no input from users. Bretz et al. (1992) suggested that research efforts be increasingly geared toward identifying the needs and preferences of users. Along similar lines, Mohrman et al. (1989) argued that the best way to ensure the acceptability of a performance appraisal system design is to let users participate in creating it.

One way to determine user preferences for proposed system alternatives would be

to conduct an experiment. The experiments conducted by Tziner, Joanis, and Murphy (2000) and Lam and Schaubroeck (1999) are examples. The random assignment of employees to different appraisal systems is not an acceptable approach in the current client organization. In essence, the current study seeks to evaluate alternative performance appraisal systems before they are implemented. Because an experimental design is not a feasible approach to answer the present research question, employee preferences for appraisal systems with different purposes must be assessed in a different way. Only a handful of studies have measured user preferences for proposed performance appraisal systems (Manshor & Kamalanabhan, 2000).

Dickinson and Zellinger (1980) studied the preferences of 86 veterinary medicine students for three different rating scale formats to be used in student evaluations of faculty. A single-item measure of preference was used: "Which form do you prefer?". Love (1981) conducted a field study of 145 police officers and 33 of their supervisors to gauge reactions to three different peer assessment methods (peer ratings, peer rankings, and peer nominations). The dependent variable was a seven-item measure of ratee reactions. The responses indicated a general negative set towards all three peer assessment methods. A MANOVA did not find any significant differences in user reactions to the three peer assessment methods. One of the risks in measuring user preferences is that users may not prefer any of the choices. The current study will thus ask respondents to rank the four appraisal systems alternative along with the choice of adopting no system.

The client organization in the current study can not rely on the results of

preference studies conducted elsewhere because the results from these studies are highly specific and not generalizable, but can use them for methodological suggestions. As a follow-up to Dickinson and Zellinger's study (1980), Wiersma and Latham (1987) conducted a field study to examine the rating scale preferences of 27 technical managers and 38 programmer-analysts in a financial institution. Wiersma and Latham (1987) were interested in preferences for three different rating scales: behavioral observation scales (BOS), behavioral expectations scales (BES), and trait scales. Each rating scale was measured with a 24-item attitude scale representing seven different criteria. A composite of all 24 items formed an eighth criterion. At the end of the questionnaire, respondents were asked to rank order the three alternative scales. This study illustrated how user preferences might be measured with both ratings and rankings in the same questionnaire. Both ratings and rankings will be incorporated into the current study.

In a field study of 88 faculty members, Farh, Werbel, and Bedeian (1988) compared a self-appraisal based performance evaluation (SABPE) with a traditional supervisory evaluation (TSE). The two systems were compared on four criteria - fairness, accuracy, comfortableness, and superiority - using a 7-point measurement scale with anchors corresponding to the criteria. For example, the anchors for the fairness criterion ranged from "Definitely less than TSE" (1) through "About the same" (4) to "Definitely more than TSE" (7). In addition, faculty members were asked to judge on a three point scale (1 = SABPE, 2 = about the same, 3 = TSE) which approach they deemed more effective for evaluating their performance. While this study is an interesting approach to preference measurement, it did not use more traditional Likert scales which are generally

recommended for attitude measurement (Patten, 1998). Likert scales assume roughly equivalent intervals between alternative responses (Devellis, 1991), thus better fitting the assumptions for tests of differences among means as proposed in this study. Also, the scales used in this study are designed to compare two choices rather than four.

Waldman (1997) conducted two field studies examining user preferences among 76 employees of a Canadian telecommunications conglomerate and 200 employees of a department of the Canadian federal government. He developed an eight-item scale (α = .86) measuring preference for a 360 degree feedback appraisal system and a four-item scale (α = .69) measuring preference for a group-based performance appraisal system. The development of separate scales for each appraisal system choice made it difficult to compare relative preferences among choices. The current study seeks to measure relative preferences among choices and so it will use the same scale to measure preferences among the four systems.

Gosselin, Werner, and Hallé (1997) surveyed 265 full-time, employed Canadian MBA students for their preferences concerning seven performance management and appraisal issues. One of the survey questions asked the respondents to rank five potential uses of appraisal information: salary increases, promotion, training, career development, and displacement (layoff). This study used ranking to encourage respondents to discriminate among items. The present study intends to use both rating and ranking of appraisal system design alternatives.

Manshor and Kamalanabhan (2000) conducted a field study of user preferences for appraisal system design alternatives, including: frequency of appraisal, frequency of

feedback, source of feedback, and use of appraisal information. They surveyed 52 raters and 122 ratees in Malaysian telecommunications companies, comparing the preferences of raters and ratees on the system design alternatives. For example, respondents ranked five potential uses for appraisal information: salary increment, promotion, displacement, training, and career development. Significant differences in mean rankings were found between raters and ratees for three of the five appraisal uses, all three of which were administrative uses. Ratees ranked salary increment and promotional uses significantly higher ($p < .05$) than raters. Raters ranked use for displacement significantly higher than ratees. No significant differences were found between raters and ratees on training and career development uses. Both raters and ratees gave high rankings to these two developmental uses. Because of differences in cultures, these findings may not generalize to an American sample, and thus to the client organization. This study also illustrates the use of ranking of alternatives as proposed in the current study.

Relationship of Appraisal Purpose and User Preferences

The current study seeks to gauge user preferences for appraisal systems with different purposes. Boswell and Boudreau (1997) suggested that employee attitudes may vary depending on perceptions of how performance appraisal is used. While there are no studies of user preferences for proposed appraisal systems with different purposes, there are a handful studies of user satisfaction with appraisal systems having different purposes. It seems reasonable to assume that user acceptance of existing appraisal systems is indicative of user preferences for proposed appraisal systems. According to Dorfman et al. (1986), the administrative and developmental purposes of performance appraisals may

differentially affect ratee satisfaction. Bretz et al. (1992) concluded that research is only beginning to address how appraisal purpose affects ratee reactions to appraisal systems.

Waldman, Bass, and Einstein (1987) reported the results of a factor analysis of questionnaire items pertaining to performance appraisal satisfaction. Three factors emerged with eigenvalues above 1.0, accounting for 86% of the common variance. The factors were: (1) satisfaction with reward outcomes; (2) satisfaction with current performance; and (3) satisfaction with future development. The satisfaction with reward outcomes factor consisted of four items ($\alpha = .75$) and included items such as, "In general, how satisfied are you with the employee appraisal system as a means of allocating organizational rewards?". The satisfaction with current performance factor consisted of four items ($\alpha = .71$) and included items such as, "How frequently does your supervisor discuss how well your are doing with regard to your major performance factors?". The satisfaction with future development factor consisted of four items ($\alpha = .83$) and included items such as, "To what extent do you feel that the employee appraisal system is used to provide feedback toward future professional development?". Waldman et al.'s (1987) factor analysis supports the idea that appraisal purpose may differentially affect user acceptance of appraisal systems.

Several studies suggest that satisfaction with appraisal systems would be enhanced by having a developmental purpose. McEvoy (1990) surveyed 128 public sector managers in five organizations to gauge their approval of subordinate appraisal of managers overall. He also measured approval of subordinate appraisal of managers for developmental and administrative purposes separately. Approval of subordinate appraisal

of managers dropped significantly as it began to count toward pay and promotions.

Pooyan and Eberhardt (1989) conducted a field study of 408 employees and 257 of their supervisors in a state university. They found that discussion of non-supervisory ratees' career and personal development during the appraisal interview had a small, but positive impact on satisfaction with the appraisal session. In a longitudinal field study of 417 employees and 391 of their supervisors in a multi-industry corporation, Nathan, Mohrman, and Milliman (1991) examined the effect of career discussion in the appraisal session on ratee satisfaction. They found a significant correlation between discussion of career issues and ratee satisfaction ($r = .42$, $p < .05$). In a hierarchical regression predicting satisfaction with the appraisal, career discussion in the appraisal session was a significant predictor ($\Delta R^2 = .08$, $p < .001$). Because career issues reflect a developmental purpose, these studies suggest a preference for appraisal systems with a developmental purpose.

Boswell and Boudreau (1997) conducted a field study of 139 employees in a production tool facility. The dependent variable was satisfaction with the appraisal. Two scales were created for perceived administrative use and perceived developmental use (α = .70 and .77 respectively). Administrative uses included salary administration, promotion decisions, and termination decisions. Developmental uses included providing performance feedback, identification of individual training needs, and determination of transfers and assignments. There was a significant positive effect of perceived development use on performance appraisal satisfaction ($\beta = .26$, $t = 3.52$, $p < .01$). The effect of perceived administrative use of appraisal on satisfaction was non-significant. This study also suggests a preference for appraisal systems with a developmental as opposed to an

administrative purpose.

Other studies suggest that an administrative purpose enhances satisfaction with an appraisal system. Fedor and Bettenhausen (1989) conducted a quasi-experimental study with 208 undergraduate students using a manipulation similar to Farh, Cannella, and Bedeian (1991). The students were assigned to course sections where peer evaluations were used either for grading or feedback purposes. The dependent variable was user acceptance. Students reacted more favorably to the peer evaluation system when the peer evaluations counted toward their participation grade.

In a field study of 121 supervisory-employee pairs in a university, Dorfman et al. (1986) found that discussion of pay and advancement in the appraisal session was associated with higher levels of employee satisfaction. Because they controlled for the ratee's performance level, they concluded that a discussion of administrative consequences had effects on employee satisfaction beyond those related to performance level (see discussion of the effect of ratee rating on satisfaction below). In a field study of 102 exempt textile employees, Giles and Mossholder (1990) found that a two-item salary linkage factor significantly predicted satisfaction with the performance appraisal system ($\Delta R^2 = .10$, $p < .001$), but not satisfaction with the performance appraisal session. Roberts (1992) found that the link between performance and personnel decision making explained 7% of the variance in perceived performance appraisal system effectiveness. These studies suggests a modest preference for appraisal systems with an administrative purpose.

Studies of the relationship between appraisal purpose and user acceptance report

mixed findings. Some studies find that a developmental purpose enhances user acceptance while other studies find that an administrative purpose enhances user acceptance. None of these studies measured user preferences for appraisal systems with different purposes. Assuming that user satisfaction is indicative of user preferences, the studies reviewed do not clearly suggest a preference for either an administrative or a developmental purpose.

Relationship of Appraisal Characteristics and User Preferences

In general, research suggests that allowing employees to participate in the rating process is associated with positive employee reactions to the appraisal (e.g., Dipboye & de Pontbriand, 1981; Giles & Mossholder, 1990; Landy et al., 1978; Nathan et al., 1991; Pooyan & Eberhardt, 1989; Roberts, 1994; Tharenou, 1995). Cawley et al.'s (1998) meta analysis of 27 field studies investigating the effects of participation on employee reactions to performance appraisal found that appraisal participation was strongly related to satisfaction ($\rho = .64$). In a field study of 65 hospital employees, Silverman and Wexley (1984) found that employees who participated in the development of the rating scale were more satisfied with the appraisal session and more willing to improve their performance than were non-participants ($t = 2.92$, $p < .01$). Preference polling, as proposed in the present study, could be viewed as a form of employee participation in the design of an appraisal system. It is possible that the mere act of conducting a preference poll could influence employee attitudes towards appraisal systems in a positive direction.

Landy et al. (1978) conducted a field study of 711 exempt employees in a manufacturing company. They identified four significant predictors of perceived fairness and accuracy of performance appraisals including plans developed with supervisor for

eliminating weaknesses. The development of plans to eliminate weaknesses reflects a developmental purpose. This study thus suggests a preference for appraisal systems with a developmental purpose.

In a field study of 165 bank tellers from three banks, Dobbins Cardy, and Platz-Vieno (1990) found that participation, development of action plans, and existence of a formal rater training program were positively correlated with appraisal satisfaction. The participation in the development of action plans identified by Dobbins et al. (1990) would seem to be part of an appraisal system intended for developmental versus administrative purposes. Dobbins et al.'s (1990) study also would suggest a preference for appraisal systems with a developmental purpose.

As cited in Hedge and Borman (1995) and Hedge and Teachout (2000), Kavanagh, Hedge and colleagues found several attitudes toward the appraisal system were significant predictors of appraisal acceptability across studies. These included attitudes about whether the appraisal system allowed raters to distinguish between ratees' proficiencies. Distinguishing between ratees' proficiencies involves between-person comparisons associated with an administrative appraisal purpose (see discussion of the differences between administrative and developmental appraisals below). In contrast to the studies referenced above, the work of Kavanagh, Hedge and colleagues would suggest a preference for appraisal systems with an administrative purpose.

Using hierarchal regression, Hedge and Teachout (2000) found that six factors predicted acceptability to job incumbents conducting self and peer ratings, including consideration of situational constraints. For supervisory raters, six factors were identified

as predictors of acceptability, including consideration of situational constraints. Because of the need to make attributions in an appraisal conducted for administrative purposes, consideration of situational constraints is more important for administrative versus developmental appraisal purposes. This study also suggests a preference for appraisal systems with an administrative purpose.

Studies of appraisal system characteristics do not provide a clear prediction of preferences for appraisal systems with developmental or administrative purposes.

Relationship of Ratee Rating to User Preferences

There are no studies of the relationship of ratings most recently received by ratees and appraisal system preferences. Studies of the relationship of ratee ratings to user satisfaction should be indicative of the relationship of ratee ratings to user preferences. In a field study of 474 exempt employees of a research and development organization, Dipboye and de Pontbriand (1981) found the level of the most recent evaluation accounted for 25% of the variance in appraisal satisfaction. A number of studies report similar findings (Dobbins et al, 1990; Dobbins, Platz, & Houston, 1993; Dorfman et al., 1986; Evans & McShane, 1988; Klein & Snell, 1994; McEvoy & Buller, 1987; Nathan et al., 1991; Pearce & Porter, 1986; Russell & Goode, 1988). While most research has found a significant relationship between past ratings and appraisal satisfaction, a few studies have not (Cederblom & Lounsbury, 1980; Landy, Barnes-Farrell, & Cleveland, 1980).

The present study seeks to determine the extent to which ratees' expected performance ratings influence their preference for different appraisal systems. For

example, do ratees with lower expected ratings prefer developmental appraisals while those with higher expected ratings prefer administrative appraisals? Pearce and Porter (1986) found that subordinates can accurately report their most recent appraisals. Russell and Goode (1988) found a significant correlation between actual ratings and manager's recall of their own ratings ($r = .44$, $p < .001$). The correlation was not as high as expected due to an upward bias in recalled ratings. Ilgen, Peterson, Martin, and Boeschen (1981) found that subordinates reported their overall performance as being significantly higher than did their supervisors ($t = -2.02$, $p < .01$). While an upward bias in expected ratings is likely, measures of association with appraisal system alternatives should not be affected as long as most respondents overestimate their expected rating and there is no ceiling effect.

Relationship of Ratee Demographics and User Preferences

The present study is interested in differences in reactions to proposed performance appraisal systems among exempt professional versus non-exempt biweekly staff. Several studies suggest more educated and higher status ratees have less positive reactions to appraisal systems (Burke, Deszca, & Weitzel, 1982; Ilgen et al., 1981; McEvoy & Buller, 1987). Exempt professional positions are higher paying and have higher educational requirements and thus have higher status. Because exempt professionals have higher status, they could be expected to have less favorable reactions to the proposed performance appraisal systems. However, many of the professional staff members are also supervisory raters.

Several studies have found that rater satisfaction with appraisal systems is more positive than that of ratees (Ilgen et al., 1981; Mount, 1983; Mount, 1984; Pooyan &

Eberhardt, 1986). Hedge and Teachout (2000) found that factors influencing user acceptance of appraisal systems differed between raters and ratees. Roberts (1992) conducted a factor analysis of 22 items related to user acceptance, appraisal fairness, and appraisal system effectiveness. His analysis indicated that managerial and employee acceptance were separate and distinct constructs.

In a field study of 94 employees and 34 supervisors in three banking institutions, Williams and Levy (2000) found that supervisors were more satisfied as ratees with the appraisal system than were employees. They also found that organization level was significantly associated with both appraisal satisfaction ($r = .62$, $p < .01$) and perceived procedural fairness ($r = .52$, $p < .01$). Both of these relationships were completely mediated by perceived system knowledge. These results suggested that the effect of organizational level on employee appraisal satisfaction was significant only through its effect on employee's knowledge and understanding of the appraisal system. Of course, higher level managers who were promoted from within may have a more positive view of the appraisal system that contributed to their promotion. Williams and Levy (2000) suggested that organizations may be able to manage user acceptance of appraisal systems by increasing levels of perceived appraisal system knowledge.

The present study is interested in differences in reactions to proposed performance appraisal systems among academic supervisors who are raters only, staff supervisors who are both raters and ratees, and non-supervisory staff who are ratees only. Research would suggest that different reactions can be expected from individuals in these different roles. Raters are likely to be more positive about the proposed appraisal systems. By educating

employees about appraisal system alternatives, an employee preference poll might enhance perceived appraisal system knowledge and thus ratee acceptance.

Designing Appraisal Systems for Different Purposes

Landy and Farr (1980) proposed a component model of performance appraisal systems. Their component model consisted of system inputs, processes, and outcomes. System inputs included: (1) roles (i.e., rater and ratee); (2) vehicle (e.g., rating instrument); and (3) context (e.g., the nature of the organization, position characteristics, and appraisal purpose). These system inputs interacted to shape the appraisal process (e.g., rater training, rater cognitive processes, rater behavior, ratee participation). The appraisal process then determined the system outcomes (e.g., the psychometric properties of the ratings, user acceptance). Landy and Farr (1980) stated that the purpose of the appraisal is of central importance: it influences the type of rating instrument used as well as the rating process. Murphy and Cleveland (1995) proposed that the purpose of appraisal influences the ratings and the effectiveness of the system.

Effect of Appraisal Purpose on Rater Cognitive Processes

A number of researchers have investigated the effect of appraisal purpose on cognitive processes (Murphy & Cleveland, 1995). DeNisi, Cafferty, and Meglino (1984) proposed a cognitive model of the performance appraisal process which included appraisal purpose. They theorized that the purpose of the appraisal can determine the types of information sought and the categories used by the rater for encoding and storage in memory. Raters can seek three types of information: (1) consensus (how others perform the same task); (2) distinctiveness (how the ratee performs other job tasks); and

consistency (how the ratee has performed this particular task in the past) (DeNisi et al., 1984). When raters make comparative decisions (i.e., for administrative purposes), they seek consensus information. When providing feedback about strengths and weaknesses (i.e., for developmental purposes), raters seek distinctiveness information.

Raters search for different information depending on the purpose of the appraisal (Williams, DeNisi, Blencoe, and Cafferty, 1985). Raters were found to search for more comparative (i.e., consensus) information when they had to select one of the ratees for some treatment. In reviewing the implications of this study, Peters and DeNisi (1990) explained that raters required more information overall and more comparative information when they were required to make designation decisions, particularly for administrative purposes. The type of information sought by raters thus differs depending upon the purpose of the appraisal system, and this affects the appraisal system design.

As is discussed below, ratings are often used for both developmental and administrative purposes simultaneously. The current study seeks to compare preferences for appraisal systems with single versus multiple purposes. It has been shown that cognitive processes are different depending on the performance appraisal purpose (Bernardin & Beatty, 1984; Boswell & Boudreau, 1999). Murphy and Cleveland (1995) suggest that if the rater has one purpose in mind and then is asked to use the appraisal for an alternative purpose, the rater will have difficulty processing the information for that alternative purpose. If raters try to keep multiple purposes in mind (which is the most typical situation), there is a risk of failing to provide useful information for any of the purposes (Murphy & Cleveland, 1995). McIntyre, Smith and Hassett (1984) speculated

that rating purpose may have more of an emotional effect on raters than a cognitive effect. There is a possibility that raters prefer appraisal systems with a single rather than multiple purposes because of these difficulties.

Appraisal Purpose is Not a Unitary Construct

Appraisal purpose is not a unitary construct (Ilgen, Barnes-Farrell, & McKellin, 1993). Different researchers have used different appraisal purpose categories, particularly in laboratory experiments. Bernardin and Beatty (1992) used three different purposes in their laboratory experiment: merit pay, training and development, and retention. In laboratory experiments, Farh and Werbel (1986) studied the effect of grading and research purposes on self-appraisals and Farh et al. (1991) studied the effect of grading and feedback purposes on peer ratings. In their laboratory experiment, McIntyre et al. (1984) manipulated three purpose instructions: hiring, course improvement, and research only. In their laboratory experiment, Murphy, Balzer, Kellam, and Armstrong (1984) manipulated two purpose instructions: research only and decisions about future assistanceships. Shore, Adams, and Tashian (1998) manipulated two purposes in their laboratory experiment: feedback for development and decisions about future assistanceships. Dobbins, Cardy and Truxillo (1988, study 1) used experimental and administrative purposes in their laboratory experiment. Jawahar and Stone (1997) used merit pay and training purposes in their laboratory experiment.

Field studies tend to assess purposes more relevant to human resources practice. Bernardin, Orban, and Carlyle (1981) used two rating purposes in their field study: personnel decision making and feedback only. Harris, Smith and Champagne (1995)

compared research ratings with administrative ratings from personnel files. Clayton and Ayres' (1996) case study of an Australian government agency described two different appraisal systems: one for managers whose purpose was to support administrative decisions and one for staff whose purpose was personal development.

Cleveland, Murphy, & Williams (1989) surveyed 106 members of Division 14 (Society of Industrial-Organizational Psychology) of the American Psychological Association, asking them to rate the impact of 20 separate uses of performance appraisal. Four factors were hypothesized to fit the data and were found to be an acceptable fit using confirmatory factor analysis. The four factors were: (1) between-person comparisons; (2) within-person comparisons; (3) systems maintenance uses; and (4) documentation uses. The between-persons factor included: salary administration, promotion, retention or termination, recognition of individual performance, layoffs, and identification of poor performers. The within-persons factor included: identification of individual training needs, performance feedback, determining transfers and assignments, and identification of individual strengths and weaknesses. The system maintenance factor included: personnel planning, determination of organizational training needs, evaluation of goal achievement, assistance in goal identification, evaluation of personnel systems, and identification of organizational development needs. The documentation factor included: criteria for validation research, documenting personnel decisions, and meeting legal requirements. All four factors were positively correlated. Cleveland et al. (1989) concluded that these correlations are one indication that organizations tend to use performance appraisal for a variety of purposes rather than a single purpose. Nearly 70% of the respondents had mean

ratings above the scale midpoint (4 = moderate impact) for both the between-persons and within persons factors (Cleveland et al., 1989). These two factors correspond to administrative and developmental purposes. Gosselin et al. (1997) observed that the systems maintenance and documentation appraisal purpose factors are of more interest to the organization than to raters or ratees.

Two Most Common Purposes in Organizations: Administrative and Developmental

While a number of different definitions of appraisal purpose have been used in research, most management theorists argue that the two most prevalent purposes are administrative and developmental (Cleveland et al, 1989; Dorfman et al., 1986; Gosselin et al., 1997; Harvey, 1995; Ilgen & Feldman, 1983; Jawahar & Williams, 1997; Latham & Wexley, 1993; Mohrman et al., 1989; Murphy & Cleveland, 1995, Thomas & Bretz, 1994). Beer (1981) defined the administrative appraisal as, "seeking information from individuals on which to base rewards and make personnel decisions" (p 27). He described its goals as including: (1) to provide feedback to ratees so that they know where they stand; (2) to develop valid data for pay and promotion decisions; and (3) to help the rater in making discharge and retention decisions. Latham and Wexley (1993) described administrative appraisals as serving personnel decisions such as promotions, transfers, demotions, layoffs, terminations, salary increases, bonuses, etc. Murphy and Cleveland (1995) described administrative appraisals as a decision aid for deciding who should be promoted, terminated, given a raise, and so forth. Boswell and Boudreau (1999) defined evaluation (i.e., administrative purpose) as "comparing an individual's performance to a set standard, other organizational members, or the individual's previous performance" (p 6).

They noted that evaluation supports human resource activities such as salary administration, promotion and termination decisions, and identification and/or recognition of good or bad performance.

Beer (1981) defined the developmental appraisal as, "seeking the development of individuals through counseling, coaching, and career planning" (p 27). He described its goals as including: (1) to counsel and coach ratees so that they improve their performance and develop future potential; (2) to develop ratee organizational commitment through discussion of career development; (3) to motivate ratees through recognition and support; (4) to strengthen rater-ratee relations; and (5) to diagnose individual and organizational problems. Latham and Wexley (1993) described developmental appraisals as enhancing a ratee's abilities and motivation. Murphy and Cleveland (1995) described developmental appraisals as providing feedback to enhance future performance. Boswell and Boudreau (1999) defined development as "any effort concerned with enriching attitudes, experiences, and skills which improves the effectiveness of employees" (p 6). They offered the following examples of developmental performance appraisal use: identifying an individual's strengths and weaknesses, setting goals, and identifying training needs.

Latham and Wexley (1993) commented that the administrative and developmental purposes of appraisal are interrelated: a ratee's abilities and motivation will affect subsequent administrative decisions, and administrative decisions affect a ratee's subsequent abilities and motivation. While these two purposes are interrelated, they necessitate different appraisal system design choices (Latham & Wexley, 1993).

Differences Between Administrative and Developmental Appraisals

An appraisal system designed for administrative purposes differs significantly from one designed for developmental purposes. These differences manifest themselves along a number of dimensions:

1. Focus of comparison. Administrative appraisals focus on between-person comparisons while developmental appraisals focus on within-person comparisons (Cleveland et al., 1989; Gosselin et al., 1997; Ostroff, 1993).

2. Rater information sought. In the administrative appraisal, the rater seeks consensus information while in the developmental appraisal the rater seeks distinctiveness information (DeNisi et al., 1984).

3. Basis of comparison. Administrative appraisals are more likely to be norm-referenced while developmental appraisals are more likely to be criterion-referenced. Administrative appraisals are based on the ratee's standing relative to relevant others (Bernardin et al, 1981) while developmental appraisals are based on the ratee's performance compared to standards.

4. Method of comparison. Administrative appraisals tend to use ranking methods or forced distributions while developmental appraisals use rating methods. Administrative appraisals seek to measure performance compared to relevant others while developmental appraisals seek to measure performance of a ratee on specific work dimensions. According to Mohrman et al. (1989) forced-choice and comparison methods of appraisal are highly useful for administrative appraisals, but are low on usefulness for developmental appraisals. In contrast, critical

incident methods of appraisal are highly useful for developmental appraisals, but are low on usefulness for administrative appraisals (Mohrman et al., 1989). Supervisors tend to dislike forced choice methods.

5. Source of ratings: Administrative appraisals are less likely to use peers (Farh et al., 1991) and subordinates as ratings sources than are developmental appraisals (Murphy & Cleveland, 1995). Self-appraisals are best suited for developmental rather than administrative purposes (Bretz et al., 1992; Farh & Werbel, 1986).

6. Type of rating(s). Administrative appraisals require a single rating (e.g., a global or overall rating or the sum of dimensional ratings) while developmental appraisals require dimensional ratings (Cardy & Dobbins, 1994; Mohrman et al., 1989). The global (or overall) rating from an administrative appraisal is used as the basis for making personnel decisions. The dimensional ratings from a developmental appraisal are used to identify areas of individual performance needing improvement (Bernardin & Beatty, 1984). In two laboratory experiments involving undergraduates, Cardy and Sutton (1996) found that rankings were more accurate than ratings when the judgment mode was global rather than dimensional.

7. Focus of feedback. An administrative appraisal provides feedback about comparative performance and relative accomplishments (Mohrman et al., 1989). A developmental appraisal provides feedback on strengths and weaknesses (Mohrman et al., 1989) and areas to improve performance (Boswell & Boudreau, 1999). Developmental feedback includes the way the work is performed and how it could be improved.

8. Role of the rater. In administrative appraisals, the rater plays the role of a judge while in developmental appraisals the rater plays the role of a coach (Ilgen & Feldman, 1983; Meyer, 1991; Meyer, Kay, & French, 1965; Sashkin, 1981).

9. Role of the ratee. In an administrative appraisal, the ratee seeks to withhold negative information which could be used in decisions that adversely affect the ratee (Mohrman et al., 1989). In a developmental appraisal, the ratee seeks to seeks to identify areas of weakness needing coaching and development (Mohrman et al., 1989).

10. Appraisal session content. An administrative appraisal session typically discusses outcomes and results and whether they can be attributed to the ratee (i.e., ratee traits) or the situation (i.e., circumstances beyond the ratee's control). Ilgen and Feldman (1983) theorized that raters must decide if an event or behavior covaries with the ratee or the situation. According to Bernardin & Beatty (1984), raters have a tendency to make attributions, particularly when they must make recommendations (e.g., pay and promotions). A developmental appraisal session typically discusses ratee strengths and weaknesses and developmental actions that could improve performance.

11. Appraisal session outcome. The outcome of an administrative appraisal is information needed to make personnel decisions (e.g., pay increase, promotion). The outcome of a developmental appraisal session is often an individual developmental action plan.

12. Time perspective. Administrative appraisals look primarily at past

performance over the rating period while developmental appraisals look at improving future performance (Cummings, 1976; Mohrman et al., 1989; Prince & Lawler, 1986).

13. Appraisal frequency. Administrative appraisals are typically conducted annually in conjunction with pay allocation decisions while developmental appraisals should be conducted more often throughout the year (Boswell & Boudreau, 1999; Meyer et al., 1965). Latham and Wexley (1993) commented that an athletic coach would be fired if appraisals were only done at the end of the season. They concluded that coaching must be provided on an on-going basis.

14. Appraisal timing. Administrative appraisals tend to be conducted on a common date (i.e., focal point) while developmental appraisals tend to be conducted on individually staggered dates (Mohrman et al., 1989). A common date facilitates comparisons between ratees. Developmental appraisals tend to require lengthy discussions that are better spread over time.

These many differences between administrative and developmental appraisals are summarized in table 1 below.

Table 1

Differences Between Administrative and Developmental Appraisals

Dimension	Administrative Appraisal	Developmental Appraisal
Focus of comparison	between-persons	within-persons
Rater information sought	consensus (how others perform)	distinctiveness (how ratee performs other tasks)
Basis of comparison	norm-referenced	criterion-referenced
Method of comparison	ranking	rating
Source of ratings	rater	rater, ratee, peers, subordinates
Type of rating(s)	global (overall)	dimensional
Focus of feedback	comparative performance; relative accomplishments	areas to improve performance
Role of the rater	judge	coach
Role of the ratee	withhold negative information	identify weaknesses needing development
Appraisal session content	outcomes and results; person/situation attribution	strengths and weaknesses; development needs/actions
Appraisal session output	information for decisions	developmental plan
Time perspective	past	future
Appraisal frequency	annual	throughout the year
Appraisal timing	focal point (common date)	individually staggered dates

Conflict Between Administrative and Developmental Appraisals

A number of management theorists have concluded that the administrative and developmental purposes of performance appraisal are incompatible (e.g., Beer, 1981; Boswell & Boudreau, 1999; Harvey, 1995; Sashkin, 1981). As described above, administrative appraisals rely on between-person comparisons of global rankings while developmental appraisals rely on within-person comparisons of dimensional ratings. Between-person global rankings can lose their connection to absolute levels of performance and do not tell the ratee specifically what he or she needs to do to improve performance (Mohrman et al., 1989). Within-person dimensional ratings do not provide the information needed for administrative decision making (Mohrman et al., 1989). When an appraisal system is used for both administrative and developmental purposes, both purposes are likely to suffer (Murphy & Cleveland, 1995). In an experimental study, Jelley and Goffin (2001) found that enhancing rating accuracy for developmental purposes reduced accuracy for administrative purposes.

Administrative appraisals require raters to make judgments affecting ratees' futures (Beer, 1981). When raters communicate these judgments, they are often called upon to justify their appraisal, as organizational rewards are often at stake. This can result in an adversarial relationship, faulty listening, and low trust which is not conducive to coaching and development (Beer, 1981; Longnecker, Sims, & Gioia, 1987). For example, a ratee could successfully complete all agreed upon developmental actions and yet find his or her relative standing in an administrative appraisal unchanged. Meyer (1991) observed that the ratee's defensive reactions are so common, and the ego involvement in the salary

discussion so powerful, that attempts to coach the ratee are almost futile. Harvey (1995) observed that salary administration involves the allocation of scarce increase dollars based on evaluation, rating, and ranking. Sashkin (1981) described pay allocation as a zero-sum exercise requiring peer comparisons. The more severe the perceived administrative consequences of a negative rating, the greater the incentive for the rater to be lenient (Farh et al., 1991). Longnecker et al. (1987) asserted that pay linkages increase the likelihood that ratings will be manipulated.

Developmental appraisals cast the rater in the role of coach, listening to ratee problems and helping the ratee identify and overcome weaknesses (Meyer et al., 1965). Ilgen and Feldman (1983) referred to the "coaching functions" of appraisal systems as providing feedback and developmental support. Harvey (1995) commented that the focus of ratee development is coaching and continuous improvement. Different communication processes are required for administrative versus developmental appraisals (Beer, 1981). Coaching requires the rater to be non-threatening (Ilgen & Feldman, 1983). The rater is in the incompatible position of being both judge and helper when administrative and developmental purposes are combined (Beer, 1981; Meyer et al, 1965). The rater roles as an accurate observer/recorder and as an accurate evaluator/judge also have different cognitive task requirements (Ilgen et al., 1993).

Beer (1981) observed that giving negative feedback to a ratee makes it difficult for the rater to maintain positive relationships with the ratee. A rater's performance typically requires maintaining the commitment of ratees. When appraisals are used for administrative purposes (e.g., pay and promotion), raters' beliefs about other raters affect

how they evaluate their ratees. Relative comparisons become important in the allocation of scarce resources. If a rater believes other raters are inflating their ratings, then that rater is inclined to also inflate his or her ratings (Bernardin & Orban, 1990).

Ratings collected for multiple purposes may be different from ratings collected for a single purpose (Cleveland et al., 1989; Milkovich & Wigdor, 1991). Cleveland et al. (1989) argued that it is unlikely that the same performance appraisal will provide valid information about both between- and within-person distinctions. Accuracy in distinguishing between individuals is largely independent of accuracy in distinguishing an individual's strengths from his or her weaknesses (Murphy, Garcia, Kerkar, Martin, & Balzer, 1982). Williams et al. (1985) showed that acquiring information for one type of decision may hinder one's ability to make other types of decisions with that information. In traditional rating systems requiring the rater to focus on multiple purposes, the rater may focus on only one purpose, disregarding the others (Longnecker et al., 1987). Because the need to make administrative decisions is important to organizations, development is the function likely to be disregarded (Harvey, 1995; Roberts, 1998). The rater can spend so much time explaining and justifying administrative appraisals that discussion of how the ratee can grow and develop gets squeezed out (Latham & Wexley, 1993).

Clayton and Ayres (1996) reported a case study of an Australian government agency that implemented an appraisal system for administrative purposes for its managers and an appraisal system for developmental purposes for its staff. The developmental appraisal system for staff was successful while the administrative appraisal system for

managers had minimal effect. Because the administrative system for managers was mandated, its minimal impact may be attributed to a lack of ownership rather than the purpose of the system (Clayton & Ayres, 1996).

Research examining the conflict between administrative and developmental appraisals is rare (Boswell & Boudreau, 1999). Milkovich and Wigdor (1991) asserted that actual appraisal systems used by organizations are complex because of their multiple purposes, making hypothesis testing difficult in field settings. According to Boswell and Boudreau (1999), we know little about employee reactions to combined versus separated administrative and developmental appraisals.

The first, and perhaps most frequently cited research (Boswell & Boudreau, 1999), on multiple appraisal uses is a study conducted at General Electric by Meyer et al. (1965). Based on interviews, questionnaires and observations of appraisal sessions with 92 employees, Meyer et al. (1965) concluded that separate appraisals should be held for different purposes. As a result, about half the managers in the General Electric study adopted a separate work planning and review (WP&R) system that did not include summary judgments or ratings. In comparing the WP&R system to the traditional approach, Meyer et al. (1965) found that ratees in the WP&R group expressed significantly more favorable attitudes and were more likely to have taken actions to improve their performance. Prince and Lawler (1986) challenged Meyer et al."s (1965) recommendation to split the administrative from the developmental appraisal as being either based on logic or inferred indirectly from their data. Because Meyer et al.'s (1965) study was not designed to compare split versus combined appraisals, this is a valid

challenge.

To test the conclusions of Meyer et al. (1965), Prince and Lawler (1986) conducted a field study involving 770 employees at nine sites of a multi-industry company. The dependent variables were the utility of the appraisal, satisfaction with the appraisal session, and perceived performance improvement. The independent variables included scales measuring the salary discussion focus and the career development focus of the appraisal session. They found that salary discussion had no impact or a slightly positive impact. Boswell and Boudreau (1999) commented, however, that Prince and Lawler's (1986) manipulation was limited and the measures focused only on the extent to which salary discussion was part of the appraisal session. It is thus not clear whether developmental and administrative appraisal purposes are indeed incompatible.

Appraisal System Design Alternatives

Mohrman et al. (1989) recommended that it is critical to decide the purpose of an appraisal system before proceeding with its design. They reasoned that this is analogous to determining the destination for a ship before sailing. Murphy and Cleveland (1995), however, observed that there have been relatively few empirical tests of the idea that appraisal purpose should affect the design of an appraisal system.

Some management theorists argue that there should be separate appraisal systems for administrative and developmental purposes (Beer, 1981; Harvey, 1995; Meyer, 1991; Sashkin, 1981). Using an appraisal system for multiple purposes can be ineffective because the best system design varies for each purpose (Boswell & Boudreau, 1999). The optimal system for encouraging development differs from the optimal system for

administrative purposes. Bartol and Locke (2000) reasoned that there are also procedural justice reasons for separating administrative and developmental appraisals. If new appraisal data surfaced during a dual purpose appraisal session, raters are rarely empowered to change pay allocations without first clearing them with their immediate superiors (Bartol & Locke, 2000). Sashkin (1981) asserted that the surest way to ruin an otherwise good appraisal system is to require the rater to perform both roles in one appraisal session. Harvey (1995) claimed that as long as the administrative and developmental purposes are connected, ratees will not receive accurate feedback about their strengths and developmental needs. Roberts (1998) commented that there is no consensus in the appraisal literature on whether there should be a separation of the administrative and developmental appraisals.

It is common practice for organizations to use appraisals for multiple purposes (Ilgen et al., 1993; Meyer, 1991; Mohrman et al., 1989; Murphy & Cleveland, 1995; Ostroff, 1993). Mohrman et al., (1989) reported that 459 out of 510 firms (90.0%) used appraisals to determine merit increases and 442 of these same 510 firms (86.7%) used appraisals as a basis for providing feedback. In a survey of 250 SHRM (Society for Human Resource Management) members, Smith, Hornsby, and Shirmeyer (1996) found that most companies used performance appraisal for multiple purposes. Of the 136 companies represented, 112 (82.4%) used appraisals for salary decisions, 111 (82.2%) used appraisals to define employee objectives for the coming period, and 98 (72.1%) used appraisals to identify training needs.

Ilgen and Feldman (1983) questioned whether the administrative and

developmental purposes of appraisal can or should be separated. They reasoned that the rater and ratee can not avoid considering reward implications even if they are not discussed in the appraisal session. Further, feedback in a developmental discussion has reward implications. Ilgen and Feldman (1983) concluded that "trying to separate the two promotes a benign fiction at best, hypocrisy at worst" (p 178). Prince and Lawler (1986) challenged an assumption inherent in split appraisal discussions that ratees will accept a separate pay discussion even though it is an important topic to them. For many ratees, the only valid reason for completing an appraisal is to determine pay (Roberts, 1998). Dorfman et al. (1986) found that the administrative and developmental purposes of appraisal each had unique impacts on outcome measures. Because the administrative purpose of appraisal did not completely overwhelm the developmental purpose; they asserted that their findings together with those of Prince and Lawler (1986) provide support of Ilgen and Feldman's (1983) position that the two should not be separated.

The current study envisions presenting subjects with four different appraisal system design alternatives: (1) administrative purpose only; (2) developmental purpose only; (3) combined administrative and developmental purposes conducted at the same time; and (4) separated administrative and developmental purposes conducted in parallel at different times. Subjects will be asked to rate each alternative on an appraisal system preference scale and then rank the four alternatives along with a fifth option of adopting none of the alternative appraisal systems.

Hypotheses

The common employer practice of using appraisal systems for multiple purposes

would seem to reflect a widespread preference for appraisal systems with multiple rather than single purposes. If users can not see a separation of the developmental and administrative appraisal purposes, as asserted by Ilgen and Feldman (1983), then both raters and ratees are more likely to prefer a system with both purposes conducted simultaneously rather than separated in time. The lack of consensus among the client organization's prospective users for the originally proposed system specifications may reflect a preference for a system with both purposes conducted simultaneously. The users may have seen a system with separate administrative and developmental purposes conducted at different times as requiring more effort. The prospective users may not have perceived any conflict in accomplishing administrative and developmental purposes simultaneously.

Hypothesis 1: The client organization users will prefer a combined appraisal system with both administrative and developmental purposes conducted at the same time.

Hypothesis 1a: Users will rate this alternative more highly than the alternative systems.

Hypothesis 1b: Users will rank this alternative more highly than the alternative systems or no system.

A summary of the preference rating and ranking hypotheses is shown in table 2 below.

Table 2

Summary of Preference Rating/Ranking Hypotheses

Hypothesis	Type	Choices	Predicted Preferences
H1a	Rating	administrative only	combined/same time
		developmental only	
		combined/same time	
		separated/different times	
H1b	Ranking	administrative only	combined/same time
		developmental only	
		combined/same time	
		separated/different times	
		none of these systems	

Studies of the relationship of user demographics to user acceptance would suggest significant differences in reactions could be expected based on functional classification (i.e., academic administrators, exempt professional staff, and non-exempt biweekly staff). The previous survey conducted in the client organization, however, did not reveal a significantly different pattern of responses based on functional classification. Because the more highly educated professional staff are often supervisors, the more favorable reaction associated with the supervisory role may offset the less favorable reaction associated with more education. No significant differences based on functional classification are thus expected.

Hypothesis 2: There will be no differences in preferences based on

functional classification. That is, there will be no differences in the preferences of (1) academic administrators, (2) exempt professional staff, and (3) non-exempt biweekly staff.

Studies of the relationship of user supervisory role to user acceptance would suggest significant differences can be expected based on role (i.e., rater only, both rater and ratee, and ratee only). The previous survey conducted in the client organization, however, did not reveal a significantly different pattern of responses between supervisors and non-supervisors. To the extent that the process of surveying user preferences enhances perceived system knowledge (Williams & Levy, 2000), differences in role-related reactions to appraisal systems may be diminished. No significant differences based on role are expected.

Hypothesis 3: There will be no differences in preferences based on role. That is, there will be no differences in the preferences of (1) academic supervisors who would be raters only and are ratees under a different system, (2) staff supervisors who would be both raters and ratees, and (3) non-supervisory staff who would be ratees only.

As previously reported, the most highly ranked purpose out of 13 choices in the earlier client organization survey was providing developmental feedback. Respondents to this survey also rated developmental purposes more highly than administrative or strategic purposes. Based on the previous client organization survey, it is expected that systems with a developmental purpose only will be preferred over those with an administrative purpose only.

Hypothesis 4: Given a choice of either an appraisal system with an administrative purpose only or an appraisal system with a developmental purpose only, significantly more respondents will choose an appraisal system with a developmental purpose only.

Based on studies of the relationship between ratee ratings and user acceptance, ratee's expected ratings should influence the type of appraisal system preferred. Ratees who expect lower ratings would seem more likely to prefer an appraisal system with a developmental purpose only while ratees who expect higher ratings would seem to prefer an appraisal system with an administrative purpose only. Ratees who expect higher ratings are more likely to receive the rewards associated with administrative appraisals.

Hypothesis 5: There will be a significant difference in the mean expected performance ratings of ratees who prefer an appraisal system with a developmental purpose only versus those who prefer an appraisal system with an administrative only purpose.

Consistent with hypothesis 1, it is expected that appraisal systems with multiple purposes will be preferred over those with either a developmental or administrative purpose alone. The previous client organization survey found support for an appraisal system with both administrative and developmental purposes. This indicates support among client organization users for an appraisal system with multiple purposes.

Hypothesis 6: Given a choice of either an appraisal system with a single purpose or an appraisal system with multiple purposes, significantly more respondents will choose an appraisal system with multiple purposes.

Hypothesis 6a: Given a choice of an appraisal system with an administrative purpose only or a combined appraisal system with both administrative and developmental purposes conducted at the same time, significantly more respondents will prefer a combined appraisal system with both administrative and developmental purposes conducted at the same time.

Hypothesis 6b: Given a choice of an appraisal system with a developmental purpose only or a combined appraisal system with both administrative and developmental purposes conducted at the same time, significantly more respondents will prefer a combined appraisal system with both administrative and developmental purposes conducted at the same time.

Hypothesis 6c: Given a choice of an appraisal system with an administrative purpose only or a separated appraisal system with parallel administrative and developmental purposes conducted at different times, significantly more respondents will prefer a separated appraisal system with parallel administrative and developmental purposes conducted at different times.

Hypothesis 6d: Given a choice of an appraisal system with a developmental purpose only or a separated appraisal system with parallel administrative and developmental purposes conducted at different times, significantly more respondents will prefer a separated appraisal system with parallel administrative and developmental purposes conducted at different times.

Consistent with hypothesis 1, it is expected that users will prefer a system with combined administrative and developmental purposes conducted at the same time rather

than separated appraisals conducted at different times.

Hypothesis 7: Given a choice of a combined appraisal system with both administrative and developmental purposes conducted at the same time or a separated appraisal system with parallel administrative and developmental purposes conducted but at different times, significantly more respondents will prefer a combined appraisal system with both administrative and developmental purposes conducted at the same time.

A summary of the paired comparison hypothesis are shown in table 3 below.

Table 3

Summary of Paired Comparisons Hypotheses

Hypotheses	Choice 1	Choice 2	Predicted Preference
H3	administrative	developmental	developmental
H6a	administrative	combined/same time	combined/same time
H6b	developmental	combined/same time	combined/same time
H6c	administrative	separated/different times	separated/different times
H6d	developmental	separated/different times	separated/different times
H7	combined/same time	separated/different times	combined/same time

Study 1

Because no existing scale measured user preferences for appraisal systems with

different purposes, a scale needed to be developed. The goal of study 1 was thus to develop a performance appraisal system preference scale using a student sample which would be used in a second study with the client organization prospective users. A pilot test of the appraisal system preference scale was conducted to reduce the number of scale items to a smaller subset that were unidimensional and had acceptable reliability.

Methods

Study 1 asked students to rate four alternative appraisal systems, each having different purposes.

Participants

The initial 12-item appraisal system preference scale was pilot tested on 187 client organization students: 84 undergraduate psychology students and 103 undergraduate business students. No identifying information was collected from the students.

Procedures

The appraisal system preference questionnaires were distributed to undergraduate psychology students through the Psychology Lab Office. Participating psychology students picked-up the questionnaire and an informed consent statement from the Psychology Lab office and received credit toward required enrichment points or extra credit when they returned a completed questionnaire. The researcher presented the preference questionnaire to the undergraduate business students during regularly scheduled class meetings and gave the students the option of not responding.

To prevent an ordering effect from influencing respondents' preferences, there were four different versions of the questionnaire. The four different questionnaire

evaluation results in a single rating of the ratee's performance. This single rating is used as a basis for making administrative decisions such as merit increases and promotion. The appraisal discussion looks back at how the ratee has performed over the past rating period.

2. Developmental Appraisal **Only**:

In a developmental appraisal system, the supervisor compares information about ratee performance across different work dimensions. The supervisor seeks information about how the ratee performed different tasks relative to work standards. The supervisor's evaluation results in ratings along different work dimensions. These dimensional ratings are used to identify areas of performance strength and weakness. This information is then used to prepare developmental action plans to improve the ratee's future performance in specific dimensions.

3. Combined (i.e, both administrative and developmental appraisals conducted at the **same time**):

A combined appraisal system combines the administrative and developmental appraisals, with both conducted together. The supervisor looks at performance across ratees and across tasks to produce a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work

dimensions. Both types of ratings are discussed in the same session.

4. Separated (i.e., parallel administrative and developmental appraisals conducted at **different times)**:

In a separated appraisal system, the administrative and developmental appraisals are conducted a few months apart so as to minimize any conflict in the supervisor roles of judge and coach. The supervisor looks at performance across ratees and across tasks to produce both a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. These two different types of ratings are discussed in separate sessions a few months apart.

Measures

The dependent variable in Study 1 was the mean rating on a performance appraisal system preference scale and the independent variable was the appraisal system design. The pilot appraisal system preference scale included the following items which were developed based on the literature review:

1. This system would be a source of great satisfaction.
2. This system would produce fair evaluations.
3. This system would produce accurate evaluations.
4. This system would do little to improve performance (reverse scored).

5. This system provides too few benefits to warrant the effort it requires (reverse scored).
6. This system would be a good way to assess staff performance.
7. I would be very uncomfortable with this system (reverse scored).
8. This system would be widely accepted by professional and biweekly staff.
9. This system would be my first choice.
10. This system would be my last choice (reverse scored).
11. This system would motivate staff to be more effective.
12. Evaluations produced by this system would be biased (reverse scored).

Each of the scale items had five response choices: strongly agree, agree, neutral, disagree, and strongly disagree. A 1-5 Likert scale was used for scoring items on the scale: strongly agree = 1; agree = 2; neutral = 3; disagree = 4, and strongly disagree = 5. For reverse scored items: strongly agree = 5; agree = 4, neural = 3, disagree = 2, and strongly disagree = 1. The summary rating for an alternative was the sum of the item scores divided by the number of items.

Results

The 12-item pilot test scale was first factor-analyzed to check for unidimensionality. Four factor analyses using the principal components method were conducted: one for the ratings on each of the four alternative performance appraisal systems. The 12-item scale produced a one-factor solution when it was applied to the administrative only, combined, and separated system alternatives. For the developmental only alternative, it produced a three-factor solution. Four scale items were identified that loaded on the second and third factors: items 4, 5, 8 and 10.

Four reliability analyses of the 12-item pilot test scale were next conducted: one for the ratings of each of the four alternatives. The item-total correlations for each item were examined and the internal consistency of the scale was assessed by computing coefficient alpha (DeVellis, 1991). The 12-item scale displayed high reliability when applied to the administrative only, combined, and separated alternatives: coefficient alphas of .92, 93, and .93 respectively. For the developmental only alternative, the coefficient alpha for the 12-item scale was .65. The scale items with the three lowest item-total correlations on each of the four reliability analyses were identified. Items 4 and 12 had low item-total correlations on three of the four reliability analyses. Items 5 and 8 had low item-total correlations on two of the four reliability analyses. Items 7 and 11 had low item-total correlations on one of the four reliability analyses.

The items which loaded on other factors and which had low item-total correlations were dropped from the scale, leaving five items: 1,2,3,6 and 9. Factor analyses of this 5-item scale applied to the four alternative appraisal systems produced one-factor solutions for all four alternatives. Reliability analyses of this 5-item scale applied to the administrative only, developmental only, combined, and separated system ratings produced coefficient alphas of .88, .73, .89, and .90 respectively. Several alternative 6-item scales were evaluated, but they either produced more than a one-factor solution for the developmental only alternative or did not improve reliability. The 5-item scale using items 1,2,3,6 and 9 was thus selected for use in study 2.

The researcher observed some of the business students completing the pilot questionnaires without first reading the descriptions of the four alternative systems. The

questionnaire was thus redesigned for Study 2 (see Appendix B) with an added page describing the four alternative appraisal systems to be read before ratings were requested. To prevent an ordering effect, four versions of this added page were developed as previously described.

A 1 x 4 within-subjects ANOVA applied to the student ratings of the four alternative appraisal systems using the 5-item scale found a significant difference in the average ratings (F (1,182) = 36.85, $p < .001$). The effect size of appraisal purpose on student ratings of appraisal systems was between medium and large as measured by η^2 (Weinfurt, 2000). Appraisal purpose accounted for 17% of the variance in ratings. Each of the alternatives were rated on a scale with the most positive response, "strongly agree" = 1, and the least positive response, "strongly disagree" = 5. The lower the average rating, the more preferred the alternative. The mean ratings are shown in table 4. The post hoc significance of the difference in means was tested using pairwise comparisons with a Bonferroni adjustment for multiple comparisons.

Table 4

Mean Student Ratings of Appraisal Systems

Appraisal System Alternative	Mean Student Rating	
Combined	2.41_{A}	(most preferred)
Developmental Only	2.49_{AB}	
Separated	2.64_{B}	
Administrative Only	3.07_{C}	(least preferred)

Cell means that do not share a subscript differ at $p < .05$.

As shown in table 4, although the average rating for the combined system suggested it was the most preferred, it was not significantly different from the average rating for the developmental only alternative. The average rating of the separated system was significantly different from the combined and administrative only alternatives. The average rating of the administrative only system indicated it was the least preferred alternative and was significantly different from the other three.

Study 2

Having developed a reliable and unidimensional scale in Study 1, the goal of study 2 was to measure client organization user preferences for the four different appraisal systems.

Methods

Study 2 asked respondents to first rate the four alternative appraisal systems. After rating all four alternatives, the respondents then ranked the alternatives along with the alternative of adopting none of the systems. The respondents further indicated a preference for each appraisal system individually paired with each of the three other systems. Respondents were then asked to estimate the overall performance rating they would expect to receive if an appraisal system was adopted.

Participants

The participants in Study 2 were 12 academic administrators who supervise staff employees, 31 non-exempt biweekly staff employees, and 35 exempt professional staff employees. This represents roughly one-fourth of the client organization employees who would be users of an appraisal system if one were adopted. The participants were asked

to identify their age category, length of service category, role, functional classification, and expected rating (ratees only). The age and service categories combined did not uniquely identify an individual respondent. The distribution by functional classification of those sent questionnaires was compared to the actual distribution of respondents by functional classification to determine if there are significant differences.

To assure that participant responses were anonymous and individual responses could not be traced to an individual, respondents were be asked to identify their age and length of service by category only rather than their actual age and length of service. Participants were asked to identify their age and length of service in one of five categories. The "What is your age?" question had the following choices: below 32 years; 32 to 39 years; 40 to 47 years; 48 to 52 years; and 53 years and above. The "What is your length of service" question had the following choices: 1 year or less; 2 to 3 years; 4 to 6 years; 7 to 10 years; and 11 or more years. The age and length of service brackets were developed by looking at prospective respondent age and length of service quintiles and by choosing understandable groupings. The age and length of service distributions of those sent questionnaires were compared with the respondent age and length of service distributions to determine if there were significant differences. A preliminary cross-tabulation of 229 non-exempt biweekly and exempt professional staff plus 36 academic administrators sent questionnaires by the age and service categories is shown in table 5 below:

Table 5

Cross Tabulation of Participants by Age and Years of Service

	Years of Service Completed					
Age	< 2	2 to 3	4 to 6	7 to 10	11+	Total
below 32	25	10	5	0	0	40
32 to 39	13	11	9	5	2	40
40 to 47	18	12	7	11	17	65
48 to 52	13	13	8	5	19	58
53+	6	6	10	7	33	62
Total	75	52	39	28	71	265

Potential ratees were asked to identify the performance rating they would expect to receive if a performance appraisal system was adopted that included an overall rating. The question "If a performance appraisal system was adopted that included an overall rating, what would you expect your rating to be?", included the following five response categories: outstanding (surpasses all performance standards by a wide margin); exceeds expectations (surpasses most performance standards); successful (meets all performance standards); needs improvement (falls below a few important performance standards); and unsuccessful (fails to meet a significant number of performance standards). A 1-5 Likert scale was used for scoring items on the expected performance rating question: outstanding = 5; exceeds expectations = 4; successful = 3; needs improvement = 2; and unsuccessful = 1.

Procedures

The preference poll questionnaires were mailed to participants in individually addressed campus envelopes. The survey questionnaires were accompanied by a cover letter (see Appendix C). Mailing labels were affixed to each envelope to assure that every respondent received a questionnaire and no respondent received more than one questionnaire. The questionnaires sealed in the envelopes contained no identifying marks.

To prevent an ordering effect from influencing respondents' preferences, there were four different versions of the page describing the alternative appraisal systems as discussed in Study 1. The four different questionnaire versions were randomly assigned to participants.

Participants were given up to four weeks from the date of distribution to respond to the questionnaire. To encourage participants to respond, follow-up e-mail messages were sent to those sent questionnaires encouraging them to respond. The follow-up e-mail messages were sent three weeks prior and one week prior to the deadline and offered to provide a replacement questionnaire. Respondents were asked to return the completed questionnaires to the researcher via the campus mail system.

Materials

The user questionnaire (see Appendix B) included five sections. The first section included a page describing the four alternative appraisal systems (with different four versions as noted above). The same performance appraisal system descriptions used in Study 1 were used in Study 2. The second section repeated the descriptions followed by the 5-item appraisal system preference scale for rating each of the alternatives. The third

section asked respondents to rank the four appraisal system choices along with the option of adopting none of the systems. The fourth section included six paired forced-choice questions. Each system was paired with the other three alternative appraisal systems with respondents asked to choose one system from each pair. The fifth section included the respondent demographic information described with the participants above.

Measures

The dependent variables were the mean rating on a performance appraisal system preference scale and the mean ranking of each alternative. The independent variables included the appraisal system design alternatives, role (rater only, both rater and ratee, ratee only), and functional classification (academic administrator, exempt professional staff, non-exempt biweekly staff). An additional independent variable associated with the paired forced-choice questions was the ratee's expected performance rating.

Results

A total of 274 questionnaires were sent and 78 responses were received. This represents a response rate of 28.5% and is comparable to the 28.6% response rate for the previous merit system design questionnaire. There were no significant differences in the response rates by: functional classification (χ^2 (2, $\underline{N}$=78) = 0.75, $\underline{p}$ = .69); age (χ^2 (4, $\underline{N}$=74) = 6.64, $\underline{p}$ = .16); or length of service (χ^2 (4, $\underline{N}$=75) = 5.14, $\underline{p}$ = .27). The study findings should thus generalize to all prospective client organization users based on functional classification, age, and length of service.

The items in the 5-item scale were originally embedded in a 12-item scale. Because changing the item context could affect scale reliability, the coefficient alphas for

the 5-item scale were again checked. The ratings of the administrative only, developmental only, combined, and separated appraisal systems again produced very acceptable coefficient alphas of .94, .92, .96, and .78 respectively. Factor analyses again produced one-factor solutions for the ratings of all four alternatives.

Each of the alternatives were rated on an appraisal system preference scale with the most positive response, "strongly agree" = 1, and the least positive response, "strongly disagree" = 5. The lower the average rating, the more preferred the alternative. A 1 x 4 within-subjects ANOVA found a significant difference in the average ratings of the four alternatives (F (1,75) = 16.65 $p < .001$). The effect size of appraisal purpose on prospective user ratings of appraisal systems was between medium and large as measured by η^2. Appraisal purpose accounted for 18% of the variance in ratings. The mean prospective user ratings are shown in table 6. The post hoc significance of the difference in means was tested using pairwise comparisons with a Bonferroni adjustment for multiple comparisons.

Table 6

Mean Prospective User Ratings of Appraisal Systems

Appraisal System Alternative	Mean User Rating
Developmental Only	2.50_A (most preferred)
Combined	2.77_{AB}
Separated	3.09_B
Administrative Only	3.44_C (least preferred)

Cell means that do not share a subscript differ at $p < .05$.

Although the average rating of the developmental only system suggested it was the most preferred, it was not significantly different ($p = .68$) from the average rating of the combined alternative. The average rating of the separated system was significantly different from the administrative only and developmental only alternatives. The average rating of administrative only system indicated that it was the least preferred and was significantly different from the other three alternatives.

Each of the five alternatives were ranked with the most preferred alternative ranked "1" and the least preferred alternative ranked "5". The lower the average ranking, the more preferred the alternative. A significant portion (36%) of the respondents did not rank all five alternatives, particularly the last alternative of "none of these systems." The missing rankings were inserted based on the average of the remaining unused ranks. For example, if a respondent only ranked one alternative as his or her first choice (i.e., a "1"), the remaining four alternatives were assigned a rank of 3.5.

A 1 x 5 within-subjects ANOVA found a significant difference in the average rankings of the five alternatives ($F(1,78) = 19.44$, $p < .001$). The effect size of appraisal purpose on prospective user rankings of appraisal systems was approaching large as measured by η^2. Appraisal purpose accounted for 20% of the variance in rankings. The mean prospective user rankings are shown in table 7. The post hoc significance of the difference in means was tested using pairwise comparisons with a Bonferroni adjustment for multiple comparisons. While an ANOVA assumes an interval scale and rankings are an ordinal scale, the ANOVA "is a very robust statistical procedure , and the assumptions can be violated with relatively minor effects" (Howell, 1997 p 321). Manshor and

Kamalanabhan (2000) used t-tests to compare mean rankings of ratees versus raters. Wiersma and Latham (1986) used an ANOVA to test differences in rank-ordered data.

Table 7

Mean Prospective User Rankings of Appraisal Systems

Appraisal System Alternative	Mean User Rating
Combined	2.14_{A} (most preferred)
Developmental Only	2.59_{AB}
Separated	2.88_{B}
Administrative Only	3.56_{C}
None of these systems	3.83_{C} (least preferred)

Cell means that do not share a subscript differ at $p < .05$.

The average ranking of the combined system suggested it was the most preferred, but it was not significantly different ($p = .37$) from the average ranking for the developmental only alternative. The average ranking of the separated system was significantly different from all the alternatives except the developmental only alternative. The average ranking of the administrative only system was significantly different from all the alternatives except none of the systems. The alternative of none of these systems was ranked the least preferred, but was not significantly different from the administrative only alternative.

The difference in rating preferences based on functional classification was tested with a 3 between- x 4 within-subjects ANOVA and was found not to be significant ($F(2,73) = 1.20$, $p = .307$). The difference in rating preferences based on role was tested

with a 3 between- x 4 within-subjects ANOVA and was found not to be significant (F (2,73) = .97, p = .386). A 5 between- x 4 within-subjects ANOVA found no significant differences in rating preferences based on age (F (4,69) = 1.76, p = .147) or length of service (F (4,70) = .93, p = .450).

Comparing systems with a single purpose, a developmental only versus administrative only system was preferred by 77.9% versus 22.1% of the respondents (χ^2 (1, N=77) = 24.01, p > .001).

Several tests were conducted of the difference in the portion of respondents choosing a system with multiple purposes (i.e., combined or separated) over a system with a single purpose (i.e., administrative only or developmental only) were:

1. A combined versus administrative only system was preferred by 88.0% versus 12.0% of the respondents (χ^2 (1, N=75) = 43.23, p < .001).
2. A combined versus developmental system was preferred by 58.4% versus 41.6% of the respondents (χ^2 (1, N=77) = 2.20, p = .138). This difference was not statistically significant.
3. A separated versus administrative only system was preferred by 69.3% versus 30.7% of the respondents (χ^2 (1, N=75) = 11.21, p < .001).
4. A developmental only versus a separated system was preferred by 59.2% versus 40.8% of the respondents (χ^2 (1, N=76) = 2.579, p = .108). This difference was not statistically significant.

Comparing systems with multiple purposes, a combined versus separated system was preferred by 68.5% versus 31.5% of the respondents (χ^2 (1, N=73) = 9.99, p < .002).

Respondents were asked to indicate the performance rating they would expect to receive on a five point scale ranging from 1 = "unsuccessful" to 5 = "outstanding". The difference in the average expected performance ratings of those preferring a developmental only versus an administrative only performance appraisal system was tested and found not to be significant ($t = -.46$, $p = .65$). The average expected rating of those preferring a developmental only system was 3.76 compared to an average of 3.65 for those who preferred an administrative only system (4 = "exceeds expectations"). Overall, the average expected performance rating was 3.74. There was no significant difference in the average rating expected by professional (3.85) compared to biweekly clerical (3.60) staff ($t = 1.35$, $p = .181$). Two-thirds (66.7%) of the staff expected a rating equal to or greater than "exceeds expectations".

Overall, the study found support for some hypotheses and not others. A summary of the hypotheses and associated findings appears in table 8.

Table 8

Summary of Hypotheses and Findings

Hypothesis	Test	Significant	Supported
1a	1 x 4 within-subjects ANOVA	Yes; $p < .001$	No; developmental only rated slightly, but not significantly, higher than combined
1b	1 x 5 within-subjects ANOVA	Yes; $p < .001$	Almost; rankings of combined slightly, but not significantly, higher than developmental only
1	See 1a, 1b above		No; combined not significantly different from developmental only
2	3 between- x 4 within-subjects ANOVA	No; $p = .307$	Yes; no difference in ratings based on functional classification
3	3 between x 4 within-subjects ANOVA	No; $p = .386$	Yes; no difference in ratings based on role
4	chi-square	Yes; $p < .001$	Yes; developmental only preferred over administrative only
5	t-test	No; $p = .65$	No; no difference in average expected rating based on above preferences

Table 8

Summary of Hypotheses and Findings

Hypothesis	Test	Significant	Supported
1a	1 x 4 within-subjects ANOVA	Yes; $p < .001$	No; developmental only rated slightly, but not significantly, higher than combined
1b	1 x 5 within-subjects ANOVA	Yes; $p < .001$	Almost; rankings of combined slightly, but not significantly, higher than developmental only
1	See 1a, 1b above		No; combined not significantly different from developmental only
2	3 between- x 4 within-subjects ANOVA	No; $p = .307$	Yes; no difference in ratings based on functional classification
3	3 between x 4 within-subjects ANOVA	No; $p = .386$	Yes; no difference in ratings based on role
4	chi-square	Yes; $p < .001$	Yes; developmental only preferred over administrative only
5	t-test	No; $p = .65$	No; no difference in average expected rating based on above preferences

Table 8 (Continued)

Summary of Hypotheses and Findings(continued)

6a	chi-square	Yes; $p < .001$	Yes; combined preferred over administrative only
6b	chi-square	No; $p = .138$	No; combined only slightly favored over developmental only
6c	chi-square	Yes; $p < .001$	Yes; separated preferred over administrative only
6d	chi-square	No; $p = .108$	No; developmental slightly, but not significantly, preferred over separated
6	see 6a,6b,6c, 6d above		Partially; multiple purposes preferred only when administrative only is the other choice
7	chi-square	Yes; $p < .002$	Yes; combined preferred over separated

Discussion

Implications for the Client Organization

The findings of this study can be used to develop a more acceptable performance appraisal system for the client organization. The combination of ratings, rankings, and paired forced-choice questions produced rather consistent findings. The originally proposed system specifications presented to the client organization called for a separated performance appraisal system. The lack of consensus within the client organization for

adopting the proposed system specifications can now be, at least partly, attributed to the relative lack of support for a separated performance appraisal system. While a separated system was consistently preferred over an administrative only system in the ratings, rankings, and paired force-choice questions, it was not preferred over the combined or developmental only systems.

The study findings do not clearly identify a preference for a combined over a developmental only system. The developmental only system was rated slightly higher than the combined alternative, but the difference was not statistically significant. The combined system was ranked slightly higher than the developmental only alternative, but the difference was not statistically significant. In the paired forced-choice question, more respondents (58.4%) preferred a combined system over a developmental only (41.6%) system, but the difference was not statistically significant. The study findings suggest that a combined system with more of a developmental focus would be most accepted by prospective users.

The proposed system specifications called for a separate developmental assessment to be conducted by raters a few months after an administrative rating. The system specifications should be changed to conduct the developmental assessment at the same time as the rating of job specific expectations. This would move toward a combined performance appraisal system with more of a developmental focus. Raters could then concentrate more on distinctiveness information, within-person comparisons, and criterion referenced ratings of specific work dimensions. The performance appraisal discussion could then focus more on strengths and weaknesses across work dimensions, areas

needing improvement, and developmental activities that would improve performance.

The administrative only system was consistently the least preferred alternative on the ratings, rankings, and paired force-choice questions. The proposed system specifications called for the development of divisional comparative ratings. These divisional comparative ratings were to be developed by a divisional reviewer or a divisional review committee ranking supervisory ratings. The system specifications should delete these divisional comparative ratings. This would move the system specifications further toward a combined system with more of a developmental focus. It might also address some of the complaints about the initially proposed system specifications being too burdensome.

The initial system specifications called for the divisional comparative ratings to be used in determining merit increases. Adopting merit pay without these comparative ratings would move the client organization further away from an administrative only appraisal system. Merit pay without these comparative ratings could consist of two parts: a policy of no merit increases for staff on formal written performance warnings under a progressive discipline system and merit by exception (Coens & Jenkins, 2000). In essence, this approach focuses on the two tails of the performance distribution and assumes that most performance variations in the middle are largely system driven. Merit by exception means awarding merit increases only to those who are truly outstanding (i.e., outstanding = "to stand out"). Under merit by exception, a committee would review nominations for exceptional merit increases. All committee members would need to agree that a nominated individual was outstanding before an award was granted (Coens &

Jenkins, 2000).

The initial system specifications included an overall supervisory rating of performance. The overall rating of performance should be retained and used as a factor in nominating staff for exceptional merit awards. Only 9.5% of the respondents expected an overall rating of "outstanding". Unlike traditional merit increase systems, however, the rating of one or a few raters is not sufficient to justify an exceptional merit increase. Truly outstanding performance must be widely recognized throughout the organization under a merit by exception approach. The portion of staff members expecting exceptional merit awards should thus not be excessively great. The fact that 66.7% of the client organization respondents expected an "exceeds expectations" overall rating or greater suggests that many employees would be disappointed if a more traditional merit increase system was adopted. If merit increases were based on supervisory ratings alone, as is done in a traditional merit increase system, a significant portion of staff would end-up being disappointed.

The merit by exception approach should lessen the inherent conflict between the administrative and developmental rating purposes. By not directly linking the overall performance rating to a merit increase amount, a rater can focus on improving the ratee's performance rather than justifying a merit increase amount. The rater and ratee can become allies rather than adversaries. The rater can concentrate on coaching the ratee in achieving wider organizational recognition for excellent performance, thereby qualifying for an increase under a merit by exception approach.

The expected performance ratings did not influence preferences for appraisal

systems with administrative only versus developmental only purposes. While it was hypothesized that those preferring an administrative only system would expect higher ratings, those preferring the developmental only alternative actually had slightly higher, but not significantly different, expected ratings. The preference for a combined system with a developmental focus can not be attributed to low performers seeking to escape accountability.

The response rate of academic administrators was 34.5%, not significantly different from the other functional classifications. Under any staff performance rating system, academic administrators will be raters only. If many academic administrators were too busy to respond to the preference poll, they may also be too busy to administer a performance rating system. The proposed system specifications called for raters and ratees to agree on job specific performance expectations by March 1st of each year. This activity is scheduled to occur at the same time of year when the preference poll was conducted. The low preference poll response rate may signal a low level of effort in administering an appraisal system in the client organization.

Implications for Future Preference Studies

While user acceptance may be a major concern of practicing managers in America's largest companies (Thomas & Bretz, 1994), appraisal systems typically fail because they are designed by personnel specialists with limited input from prospective users (Schellhardt, 1996). Had the client organization implemented the initially proposed performance appraisal system specifications, it potentially could have ended-up with a number of dissatisfied users. The client organization would then have typified most

employers. A survey conducted by the Society for Human Resources Management concluded that more than 90% of appraisal systems are unsuccessful (Schellhardt, 1996). A 1993 survey by Development Dimensions International of 1,149 human resources managers and others found overwhelming dissatisfaction with appraisal systems (Schellhardt, 1996).

With the benefit of 20/20 hindsight, the user organization should have first surveyed user preferences for alternative appraisal systems before surveying preferences for specific system components. Hopefully, other employers can learn from this experience. The descriptions of the four appraisal system alternatives could be used by other employers in similar user preference studies. Study 1 developed a unidimensional and reliable appraisal system preference scale which others might use to measure preferences for alternative appraisal systems.

Two client organization clerical support staff members responded saying they could not complete the survey because they did not understand the choices. For example, one respondent did not understand in an administrative only system how supervisory ratings would be pooled when a supervisor has only one or a few ratees to rate. Explaining the various alternative ways a rater's supervisor could pool administrative ratings to facilitate between-person comparisons could make the choice descriptions too complex. While there were no significant differences in the response rates among functional classifications, the response rate for clerical support staff (26.1%) was slightly less than that for professional staff (29.2%). No service-maintenance employees were included in the study 2 sample. If service-maintenance employees, who are generally not

as well educated, had been included in the sample, they may have experienced a significantly lower response rate. Preference survey designers need to be cautious in matching the complexity of the choices to the conceptual abilities of the prospective users.

As hypothesized, there were no significant differences in client organization preferences based on functional classification or role. These hypotheses were based on previous experience with client organization surveys rather than the literature review. Studies cited in the literature review found differences in reactions to appraisal systems based on respondent demographics. The dependent variable measured in these studies was primarily satisfaction rather than preferences. Perhaps respondents can share preferences for proposed appraisal systems, but have different levels of satisfaction with implemented systems based on their function or role.

The preferences of the client organization's prospective users are unique to that organization. Prospective users in other organizations will most likely have different preferences because of differences in their missions, climates, and cultures. While the findings of study 2 do not generalize beyond the client organization, together with the findings of study 1, they raise questions about employee preferences in general. Both the student sample and the client organization sample were consistent in preferring either a combined or a developmental only appraisal system. Both samples were also consistent in least preferring an administrative only system. Perhaps the widespread dissatisfaction with performance appraisal systems is related to too much emphasis on administrative purposes in the eyes of system users.

Designing appraisal systems based on user preferences may not translate into user

satisfaction with those systems. The common organizational practice of using appraisals for multiple purposes suggests that prospective users generally prefer a combined system. The widespread dissatisfaction with appraisal systems could be related to the common practice of using appraisals for both administrative and developmental purposes. Combining administrative and developmental purposes may be incompatible as argued by some management theorists.

Limitations of the Current Study

The relatively low response rate (28.5%) in study 2 could limit the generalizability of the study findings to all client organization prospective users. Respondents could differ from those who did not respond on variables other than functional classification, age, or length of service. For example, there possibly could have been significant differences in response rates from different divisions within the client organization. Perhaps those who participated in divisional discussions of the proposed system specifications better understood the choices. Adding respondent demographic variables would have increased the likelihood of uniquely identifying a respondent and may have diminished the response rate. Conducting meetings with all employees to discuss the choices before conducting the preference poll may have increased the response rate and enhanced respondent understanding of the choices.

The descriptions of the four appraisal system choices did not include estimates of how much time and effort each may require of the rater and the ratee. If such information was available, it might have dramatically affected the preferences of the prospective users. Each respondent had to make his or her own assumptions about the amount of time and

effort required. The estimates of some individuals could have been unrealistic.

A high percentage (36%) of respondents did not rank all five alternatives. The filling-in of missing values based on the average of the unused ranks may not have reflected the intent of the respondents. While this is a common approach to filling-in missing ranking values, it could have distorted the findings. The consistency of the ranking outcomes with the ratings and paired forced-choice questions suggests that no such distortion occurred.

While Study 2 provides a clearer picture of the preferences of prospective client organization users, it does not explain why they preferred one system over another. The study did not probe into the reasons for prospective user choices. We don't know if they saw a conflict in accomplishing administrative and developmental purposes simultaneously. We don't know if they saw a separated system as requiring more effort. We don't know why they saw an administrative only system as the least preferred. Knowing the reasons for prospective user preferences would aid in assessing the extent to which the findings of this study are applicable to other organizations.

Directions for Future Research

Future research could determine whether performance appraisal systems designed based on prospective user input result in greater appraisal system satisfaction than those designed by personnel specialists alone. An assumption inherent in this study is that designing an appraisal system based on user input will lead to greater user acceptance. A related research issue is whether user preferences lead to incompatible appraisal system designs. Do prospective users prefer appraisal system design features which are

incompatible (e.g., rankings and individually staggered appraisal dates)?

Studies 1 and 2 found similar patterns of preferences. Appraisal system preference polls could be conducted in other organizations, industries, or regions to determine if this pattern of preferences is widespread. Will respondents in other samples most prefer combined and developmental only appraisal systems and least prefer administrative systems? If this pattern is widespread, it would have broad implications for the design of appraisal systems.

Qualitative research techniques could be used to probe more deeply into the reasons prospective users prefer some appraisal systems over others. Understanding these reasons would have implications for the design, implementation, and communication of appraisal systems. Qualitative research techniques might also be used to determine why prospective users do not respond to preference polls. Does a relatively low response rate signal that little effort will be put into making an appraisal system work?

References

Bannister, B. D. (1986). Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. Journal of Applied Psychology, 71. 203-210.

Bartol, K. M., & Locke, E. A. (2000). Incentives and motivation. In S. L. Rynes & B. A. Gerhart (Eds), Compensation in Organizations: Current Research and Practice. (pp. 104-147). San Francisco: Jossey-Bass Inc., Publishers.

Beer, M. (1981). Performance appraisal: Dilemmas and possibilities. Organization Dynamics, 9 (1). 24-36.

Bernardin, H. J., & Beatty, R. W. (1984). Performance Appraisal: Assessing Human Behavior at Work. Boston: Kent Publishing Company.

Bernardin, H. J., & Orban, J. A. (1990). Leniency effect as a function of rating format, purpose for appraisal, and rater individual differences. Journal of Business and Psychology, 5. 197-211.

Bernardin, H. J., Orban, J. A., & Carlyle, J. J. (1981). Performance rating as a function of trust in appraisal and rater individual differences. Proceedings of the Academy of Management. 311-315.

Boswell, W. R., & Boudreau, J. W. (1997). Employee satisfaction with the performance appraisal and the appraiser: The role of perceived appraisal use. Cornell University, Center for Advanced Human Resource Studies, Working Paper 97-14.

Boswell, W. R., & Boudreau, J. W. (1999). Separating the developmental and evaluative performance appraisal uses. Cornell University, Center for Advanced Human

Resource Studies, Working Paper 99-09.

Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. Journal of Management, 1. 321-352.

Burke, R. J., Deszca, G., & Weitzel, W. (1982). Subordinate expectations and satisfaction with the performance appraisal interview. The Journal of Psychology, 111. 41-49.

Cardy, R. L., & Dobbins, G. H. (1994). Performance Appraisal: Alternative Perspectives. Cincinnati, OH: South-Western Publishing Co.

Cardy, R. L., & Sutton, C. L. (1996). Judgment type and mode: An accuracy comparison of global and dimensional rating and ranking. Paper presented at the 1996 Annual Meeting of the Midwest Decision Sciences Institute.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. Journal of Applied Psychology, 83. 615-633.

Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. Personnel Psychology, 33. 567-579.

Clayton, P., & Ayres, H. (1996). Performance appraisal or performance development? A tale of two schemes. Australian Journal of Public Administration, 55. 63-72.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74.

130-135.

Coens, T. & Jenkins, M. (2000). Abolishing Performance Appraisals. San Francisco: Brett-Koehler Publishers, Inc.

Cummings, L. L. (1976). Appraisal purpose and the nature, amount, and frequency of feedback. Washington, D.C.: U. S. Department of Health, Education and Welfare, National Institute of Education. (ERIC Document Reproduction Service No. ED136169).

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33. 360-396.

DeVellis, R. F. (1991). Scale Development: Theory and Applications. Newbury Park, CA: Sage Publications, Inc.

Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored and mixed standard scale formats. Journal of Applied Psychology, 65. 147-154.

Dipboye, R. L., & de Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. Journal of Applied Psychology, 66. 248-251.

Dobbins, G. H., Cardy, R. L., & Platz-Vieno, S. J. (1990). A contingency approach to appraisal satisfaction: An initial investigation of the joint effects of organizational variables and appraisal characteristics. Journal of Management, 16. 619-632.

Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. (1988). The effects of purpose

of appraisal on individual differences in stereotypes of women on sex differences in performance ratings: A laboratory and field study. Journal of Applied Psychology, 73. 551-558.

Dobbins, G. H., Platz, S. J., & Houston, J. (1993). Relationship between trust in appraisal and appraisal effectiveness: A field study. Journal of Business and Psychology, 7. 309- 322.

Dorfman, P. W., Stephan, W. G., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. Personnel Psychology, 39. 579-597.

Evans, E. M., & McShane, S. L. (1988). Employee perceptions of performance appraisal fairness in two organizations. Canadian Journal of Behavioural Science, 20. 177-191.

Farh, J., Cannella, A. A., Jr., & Bedeian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. Group & Organization Studies, 16. 367-386.

Farh, J., & Werbel, J. D. (1986). Effects of purpose of the appraisal and expectations of validation on self-appraisal leniency. Journal of Applied Psychology, 71. 527-529.

Farh, J., Werbel, J. D., & Bedeian, A. G. (1988). An empirical investigation of self-appraisal based performance evaluation. Personnel Psychology, 41. 141-156.

Fedor, D. B., & Bettenhausen, K. L. (1989). The impact of purpose, participant preconceptions, and rating level on the acceptance of peer evaluations. Group &

Organizational Studies, 14 (2). 182-197.

Giles, W. F., & Mossholder, K. W. (1990). Employee reactions to contextual and session components of performance appraisal. Journal of Applied Psychology, 75. 371-377.

Gosselin, A., Werner, J. M., & Hallé, N. (1997). Ratee preferences concerning performance management and appraisal. Human Resource Development Quarterly, 8. 315-333.

Harris, M. M., Smith, D. E., & Champagne, D. (1995). A field study of performance appraisal purpose: Research- versus administrative-based ratings. Personnel Psychology, 48. 151-161.

Harvey, E. L. (1995). Separating employee development from salary systems. Human Resource Professional, 8. 3-5.

Hedge, J. W., & Borman, W. C. (1995). Changing conceptions and practices in performance appraisal. In A. Howard (Ed.), The Changing Nature of Work. (pp. 451-481). San Francisco: Jossey-Bass Inc., Publishers.

Hedge, J. W., & Teachout, M. S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. Group & Organization Management, 25. 22-44.

Howell, D. C. (1997). Statistical Methods for Psychology (4th Edition). Belmont, CA: Duxbury Press, An Imprint of Wadsworth Publishing Company.

Ilgen, D. R., Barnes-Farrell, J. L., & McKellin, D. B. (1993). Performance appraisal process research in the 1980's: Has it contributed to appraisals in use?

Organization Behavior and Human Decision Processes, 54. 321-368.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. Research in Organization Behavior, 5. 141-197.

Ilgen, D. R., Peterson, R. B., Martin, B. A., & Boeschen, D. A. (1981). Supervisor and subordinate reactions to performance appraisal sessions. Organizational Behavior and Human Performance, 28. 311-330.

Jawahar, I. M., & Stone, T. H. (1997). Influence of raters' self-consciousness and appraisal purpose on leniency and accuracy of performance ratings. Psychological Reports, 80. 323-336.

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. Personnel Psychology, 50. 905-926.

Jelley, R. B., & Goffin, R. D. (2001). Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. Journal of Applied Psychology, 86. 134-144.

Jordan, J. L. (1990). Performance appraisal satisfaction and supervisors' traits. Psychological Reports, 66. 1337-1338.

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. Journal of Applied Psychology, 85. 708-723.

Klein, H. J., & Snell, S. A. (1994). The impact of interview process and context on performance appraisal interview effectiveness. Journal of Managerial Issues, 6 (2). 160-175.

Lam, S. S., & Schaubroeck, J. (1999). Total quality management and

performance appraisal: An experimental study of process versus results and group versus individual approaches. Journal of Organizational Behavior, 20. 445-457.

Landy, F. J., Barnes, J. L., & Murphy, K. R. (1978). Correlates of perceived accuracy of performance evaluation. Journal of Applied Psychology, 63. 751-754.

Landy, F. J., Barnes-Farrell, J., & Cleveland, J. N. (1980). Perceived fairness and accuracy of performance evaluation: A follow-up. Journal of Applied Psychology, 65. 355-356.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87. 72-107.

Latham, G. P., & Wexley, K. N. (1993). Increasing Productivity Through Performance Appraisal. Reading, MA: Addison-Wesley Publishing Company, Inc.

Longnecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the Mask: The politics of employee appraisal. Academy of Management Executive, 1 (3). 183-193.

Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. Journal of Applied Psychology, 66. 451-457.

Manshor, A. T., & Kamalanabhan, T. J. (2000). An examination of raters' and ratees' preferences in process and feedback in performance appraisal. Psychological Reports, 86. 203-214.

McEvoy, G. M., & Buller, P. F. (1987). User acceptance of peer appraisals in an industrial setting. Personnel Psychology, 40. 785-797.

McEvoy, G. M. (1990). Public sector managers' reactions to appraisal by subordinates. Public Personnel Management, 19 (2). 201-213.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69. 147-156.

Meyer, H. H. (1991). A solution to the performance appraisal feedback enigma. Academy of Management Executive, 5 (1). 68-76.

Meyer, H. H., Kay, E., & French, J. R., Jr. (1965). Split roles in performance appraisal. Harvard Business Review, 43. 123-129.

Milkovich, G. T., & Wigdor, A. K. (1991). Pay for Performance: Evaluating Performance Appraisal and Merit Pay. Washington, D. C.: National Academy Press.

Mohrman, A. M., Jr., Resnick-West, S. M., & Lawler, E. E., III. (1989). Designing Performance Appraisal Systems: Aligning Appraisals and Organizational Realities. San Francisco: Jossey-Bass Inc., Publishers.

Mount, M. K. (1983). Comparisons of managerial and employee satisfaction with a performance appraisal system. Personnel Psychology, 36. 99-110.

Mount, M. K. (1984). Satisfaction with a performance appraisal system and appraisal discussion. Journal of Occupational Behaviour, 5. 271-279.

Murphy, K. R., Balzer, W. K., Kellam, K. L., & Armstrong, J. G. (1984). Effects of purpose of rating on accuracy in observing teacher behavior and evaluating teacher performance. Journal of Educational Psychology, 76. 45.54.

Murphy, K. R., & Cleveland, J. N. (1995). Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives. Thousand Oaks, CA: Sage Publications, Inc.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. (1982). The relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 67. 320-325.

Nathan, B. R., Mohrman, A. M., Jr., & Milliman, J. (1991). Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: A longitudinal study. Academy of Management Journal, 34. 352-369.

Ostroff, C. (1993). Rater perceptions, satisfaction, and performance ratings. Journal of Occupational and Organizational Psychology, 66. 345-356.

Patten, M. L. (1998). Questionnaire Research. Los Angeles: Pyrczak Publishing.

Pearce, J. L., & Porter, L. W. (1986). Employee responses to formal performance appraisal feedback. Journal of Applied Psychology, 71. 211-218.

Peters, L. H., & DeNisi, A. S. (1990). An information processing role for appraisal purpose and job type in the development of an appraisal system. Journal of Managerial Issues, II (2). 160-175.

Pooyan, A., & Eberhardt, B. J. (1989). Correlates of performance appraisal satisfaction among supervisory and non-supervisory employees. Journal of Business Research, 19. 215-226.

Prince, J. B., & Lawler, E. E. III (1986). Does salary discussion hurt the developmental performance appraisal? Organizational Behavior and Human Decision Processes, 37. 357-375.

Roberts, G. E. (1992). Linkages between performance appraisal system effectiveness and rater and ratee acceptance. Review of Public Personnel Administration,

12 (3). 19-41.

Roberts, G. E. (1994). Maximizing performance appraisal system acceptance: Perspectives from municipal government personnel administrators. Public Personnel Management, 23. 525-550.

Roberts, G. E. (1998). Perspectives on enduring and emerging issues in performance appraisal. Public Personnel Management, 27. 301-321.

Russell, J. S., & Goode, D. L. (1988). An analysis of managers' reactions to their own performance appraisal feedback. Journal of Applied Psychology, 73. 63-67.

Sashkin, M. (1981). Appraising appraisal: Ten lessons from research for practice. Organizational Dynamics, 9 (1). 37-50.

Schellhardt, T. D. (1996, November 19). It's time to evaluate your work, and all involved are groaning. The Wall Street Journal. P A1:A5.

Shore, T. H., Adams, J. A., & Tashian, A. (1998). Effects of self-appraisal information, appraisal purpose, and feedback target on performance appraisal ratings. Journal of Business and Psychology, 12 (3). 283-299.

Silverman, S. B., & Wexley, K. N. (1984). Reaction of employees to performance appraisal interviews as a function of their participation in rating scale development. Personnel Psychology, 37. 703-710.

Smith, B. N., & Hornsby, J. S., & Shirmeyer, R. (1996). Current trends in performance appraisal: Examination of managerial practice. SAM Advanced Management Journal, 61 (3). 10-16.

Tharenou, P. (1995). The impact of a developmental performance appraisal

program on employee perceptions in an Australian federal agency. Group & Organization Management, 20 (3). 245-272.

Thomas, S. L., & Bretz, R. D., Jr. (1994). Research and practice in performance appraisal: Evaluating performance in America's largest companies. SAM Advanced Management Journal, 59 (2). 28-34.

Tziner, A., Joanis, C., & Murphy, K. R. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and ratee satisfaction. Group & Organization Management, 25 (2). 175-190.

Waldman, D. A., Bass, B. M., & Einstein, W. O. (1987). Leadership and outcomes of performance appraisal processes. Journal of Occupational Psychology, 60. 177-186.

Waldman, D. A. (1997). Predictors of employee preferences for multi-rater and group-based performance appraisal. Group & Organization Management, 22 (2). 264-288.

Weinfurt, K. P. (2000). Repeated measures analyses: ANOVA, MANOVA, and HLM. In L. G. Grimm and P. R. Yarnold (Eds.), Reading and Understanding More Multivariate Statistics. Washington, D.C.: American Psychological Association.

Wiersma, U., & Latham, G. P. (1986). The practicality of behavioral observation scales, behavioral expectation scales, and trait scales. Personnel Psychology, 39. 619-628.

Williams, J. R., & Levy, P. E. (2000). Investigating some neglected criteria: The influence of organization level and perceived system knowledge on appraisal reactions.

Journal of Business and Psychology, 14. 501-514.

Williams, K. J., DeNisi, A. S., Blencoe, A. G., & Cafferty, T. P. (1985). The role of performance appraisal purpose: Effects of purpose on information acquisition and utilization. Organization Behavior and Human Decision Processes, 35. 314-339.

Appendix A

Appraisal System Design Preference Poll Questionnaire

INSTRUCTIONS: Read the descriptions of the Administrative only, Developmental only, Combined, and Separated appraisal systems and then rate each on the preference scale.

Administrative Appraisal Only

In an administrative appraisal system, the supervisor compares information across ratees. The supervisor seeks information about how others have performed similar tasks. The method of comparison is based on relative standing or ranking. When a supervisor has only one or a few ratees to rate, ratings of several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The supervisor's evaluation results in a single rating of the ratee's performance. This single rating is used as a basis for making administrative decisions such as merit increases and promotion. The appraisal discussion looks back at how the ratee has performed over the past rating period.

Appraisal System Preference Scale

1. The administrative appraisal system would be a source of great satisfaction.
   ☐ Strongly agree 1-1 ☐ Agree 1-2 ☐ Neutral 1-3 ☐ Disagree 1-4 ☐ Strongly disagree 1-5
2. The administrative appraisal system would produce fair evaluations.
   ☐ Strongly agree 2-1 ☐ Agree 2-2 ☐ Neutral 2-3 ☐ Disagree 2-4 ☐ Strongly disagree 2-5
3. The administrative appraisal system would produce accurate evaluations.
   ☐ Strongly agree 3-1 ☐ Agree 3-2 ☐ Neutral 3-3 ☐ Disagree 4-4 ☐ Strongly disagree 5-5
4. The administrative appraisal system would do little to improve performance.
   ☐ Strongly agree 4-1 ☐ Agree 4-2 ☐ Neutral 4-3 ☐ Disagree 4-4 ☐ Strongly disagree 4-5
5. The administrative appraisal system provides too few benefits to warrant the effort it requires.
   ☐ Strongly agree 5-1 ☐ Agree 5-2 ☐ Neutral 5-3 ☐ Disagree 5-4 ☐ Strongly disagree 5-5
6. The administrative appraisal system would be a good way to assess staff performance.
   ☐ Strongly agree 6-1 ☐ Agree 6-2 ☐ Neutral 6-3 ☐ Disagree 6-4 ☐ Strongly disagree 6-5
7. I would be very uncomfortable with the administrative appraisal system.
   ☐ Strongly agree 7-1 ☐ Agree 7-2 ☐ Neutral 7-3 ☐ Disagree 7-4 ☐ Strongly disagree 7-5
8. The administrative appraisal system would be widely accepted by professional and biweekly staff.
   ☐ Strongly agree 8-1 ☐ Agree 8-2 ☐ Neutral 8-3 ☐ Disagree 8-4 ☐ Strongly disagree 8-5
9. The administrative appraisal system would be my first choice.
   ☐ Strongly agree 9-1 ☐ Agree 9-2 ☐ Neutral 9-3 ☐ Disagree 9-4 ☐ Strongly disagree 9-5
10. The administrative appraisal system would be my last choice.
    ☐ Strongly agree 10-1 ☐ Agree 10-2 ☐ Neutral 10-3 ☐ Disagree 10-4 ☐ Strongly disagree 10-5
11. The administrative appraisal system would motivate staff to be more effective.
    ☐ Strongly agree 11-1 ☐ Agree 11-2 ☐ Neutral 11-3 ☐ Disagree 11-4 ☐ Strongly disagree 11-5
12. Evaluations produced by the administrative appraisal system would be biased.
    ☐ Strongly agree 12-1 ☐ Agree 12-2 ☐ Neutral 12-3 ☐ Disagree 12-4 ☐ Strongly disagree 12-5

INSTRUCTIONS: Read the descriptions of the Administrative only, Developmental only, Combined, and Separated appraisal systems and then rate each on the preference scale.

Developmental Appraisal Only

In a developmental appraisal system, the supervisor compares information about ratee performance across different work dimensions. The supervisor seeks information about how the ratee performed different tasks relative to work standards. The supervisor's evaluation results in ratings along different work dimensions. These dimensional ratings are used to identify areas of performance strength and weakness. This information is then used to prepare developmental action plans to improve the ratee's future performance in specific dimensions.

Appraisal System Preference Scale

1. The developmental appraisal system would be a source of great satisfaction.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
13-1	13-2	13-3	13-4	13-5

2. The developmental appraisal system would produce fair evaluations.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
14-1	14-2	14-3	14-4	14-5

3. The developmental appraisal system would produce accurate evaluations.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
15-1	15-2	15-3	15-4	15-5

4. The developmental appraisal system would do little to improve performance.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
16-1	16-2	16-3	16-4	16-5

5. The developmental appraisal system provides too few benefits to warrant the effort it requires.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
17-1	17-2	17-3	17-4	17-5

6. The developmental appraisal system would be a good way to assess staff performance.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
18-1	18-2	18-3	18-4	18-5

7. I would be very uncomfortable with the developmental appraisal system.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
19-1	19-2	19-3	19-4	19-5

8. The developmental appraisal system would be widely accepted by professional and biweekly staff.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
20-1	20-2	20-3	20-4	20-5

9. The developmental appraisal system would be my first choice.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
21-1	21-2	21-3	21-4	21-5

10. The developmental appraisal system would be my last choice.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
22-1	22-2	22-3	22-4	22-5

11. The developmental appraisal system would motivate staff to be more effective.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
23-1	23-2	23-3	23-4	23-5

12. Evaluations produced by the developmental appraisal system would be biased.

☐ Strongly agree	☐ Agree	☐ Neutral	☐ Disagree	☐ Strongly disagree
24-1	24-2	24-3	24-4	24-5

INSTRUCTIONS: Read the descriptions of the Administrative only, Developmental only, Combined, and Separated appraisal systems and then rate each on the preference scale.

Combined Appraisal

(Both administrative and developmental appraisals conducted at the *same time*):
A combined appraisal system combines the administrative and developmental appraisals, with both conducted together. The supervisor looks at performance across ratees and across tasks to produce a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. Both types of ratings are discussed in the same session.

Appraisal System Preference Scale

1. The combined system would be a source of great satisfaction.
   ☐ Strongly agree 25-1 ☐ Agree 25-2 ☐ Neutral 25-3 ☐ Disagree 25-4 ☐ Strongly disagree 25-5
2. The combined system would produce fair evaluations.
   ☐ Strongly agree 26-1 ☐ Agree 26-2 ☐ Neutral 26-3 ☐ Disagree 26-4 ☐ Strongly disagree 26-5
3. The combined system would produce accurate evaluations.
   ☐ Strongly agree 27-1 ☐ Agree 27-2 ☐ Neutral 27-3 ☐ Disagree 27-4 ☐ Strongly disagree 27-5
4. The combined system would do little to improve performance.
   ☐ Strongly agree 28-1 ☐ Agree 28-2 ☐ Neutral 28-3 ☐ Disagree 28-4 ☐ Strongly disagree 28-5
5. The combined system provides too few benefits to warrant the effort it requires.
   ☐ Strongly agree 29-1 ☐ Agree 29-2 ☐ Neutral 29-3 ☐ Disagree 29-4 ☐ Strongly disagree 29-5
6. The combined system would be a good way to assess staff performance.
   ☐ Strongly agree 30-1 ☐ Agree 30-2 ☐ Neutral 30-3 ☐ Disagree 30-4 ☐ Strongly disagree 30-5
7. I would be very uncomfortable with the combined appraisal system.
   ☐ Strongly agree 31-1 ☐ Agree 31-2 ☐ Neutral 31-3 ☐ Disagree 31-4 ☐ Strongly disagree 31-5
8. The combined appraisal system would be widely accepted by professional and biweekly staff.
   ☐ Strongly agree 32-1 ☐ Agree 32-2 ☐ Neutral 32-3 ☐ Disagree 32-4 ☐ Strongly disagree 32-5
9. The combined appraisal system would be my first choice.
   ☐ Strongly agree 33-1 ☐ Agree 33-2 ☐ Neutral 33-3 ☐ Disagree 33-4 ☐ Strongly disagree 33-5
10. The combined appraisal system would be my last choice.
    ☐ Strongly agree 34-1 ☐ Agree 34-2 ☐ Neutral 34-3 ☐ Disagree 34-4 ☐ Strongly disagree 34-5
11. The combined appraisal system would motivate staff to be more effective.
    ☐ Strongly agree 35-1 ☐ Agree 35-2 ☐ Neutral 35-3 ☐ Disagree 35-4 ☐ Strongly disagree 35-5
12. Evaluations produced by the combined appraisal system would be biased.
    ☐ Strongly agree 36-1 ☐ Agree 36-2 ☐ Neutral 36-3 ☐ Disagree 36-4 ☐ Strongly disagree 36-5

INSTRUCTIONS: Read the descriptions of the Administrative only, Developmental only, Combined, and Separated appraisal systems and then rate each on the preference scale.

Separated Appraisal

(parallel administrative and developmental appraisals conducted at *different times*):
In a separated appraisal system, the administrative and developmental appraisals are conducted a few months apart so as to minimize any conflict in the supervisor roles of judge and coach. The supervisor looks at performance across ratees and across tasks to produce both a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. These two different types of ratings are discussed in separate sessions a few months apart.

Appraisal System Preference Scale

1. The separated system would be a source of great satisfaction.
   ☐ Strongly agree 37-1 ☐ Agree 37-2 ☐ Neutral 37-3 ☐ Disagree 37-4 ☐ Strongly disagree 37-5
2. The separated system would produce fair evaluations.
   ☐ Strongly agree 38-1 ☐ Agree 38-2 ☐ Neutral 38-3 ☐ Disagree 38-4 ☐ Strongly disagree 38-5
3. The separated system would produce accurate evaluations.
   ☐ Strongly agree 39-1 ☐ Agree 39-2 ☐ Neutral 39-3 ☐ Disagree 39-4 ☐ Strongly disagree 39-5
4. The separated system would do little to improve performance.
   ☐ Strongly agree 40-1 ☐ Agree 40-2 ☐ Neutral 40-3 ☐ Disagree 40-4 ☐ Strongly disagree 40-5
5. The separated system provides too few benefits to warrant the effort it requires.
   ☐ Strongly agree 41-1 ☐ Agree 41-2 ☐ Neutral 41-3 ☐ Disagree 41-4 ☐ Strongly disagree 41-5
6. The separated system would be a good way to assess staff performance.
   ☐ Strongly agree 42-1 ☐ Agree 42-2 ☐ Neutral 42-3 ☐ Disagree 42-4 ☐ Strongly disagree 42-5
7. I would be very uncomfortable with the separated appraisal system.
   ☐ Strongly agree 43-1 ☐ Agree 43-2 ☐ Neutral 43-3 ☐ Disagree 43-4 ☐ Strongly disagree 43-5
8. The separated appraisal system would be widely accepted by professional and biweekly staff.
   ☐ Strongly agree 44-1 ☐ Agree 44-2 ☐ Neutral 44-3 ☐ Disagree 44-4 ☐ Strongly disagree 44-5
9. The separated appraisal system would be my first choice.
   ☐ Strongly agree 45-1 ☐ Agree 45-2 ☐ Neutral 45-3 ☐ Disagree 45-4 ☐ Strongly disagree 45-5
10. The separated appraisal system would be my last choice.
    ☐ Strongly agree 46-1 ☐ Agree 46-2 ☐ Neutral 46-3 ☐ Disagree 46-4 ☐ Strongly disagree 46-5
11. The separated appraisal system would motivate staff to be more effective.
    ☐ Strongly agree 47-1 ☐ Agree 47-2 ☐ Neutral 47-3 ☐ Disagree 47-4 ☐ Strongly disagree 47-5
12. Evaluations produced by the separated appraisal system would be biased.
    ☐ Strongly agree 48-1 ☐ Agree 48-2 ☐ Neutral 48-3 ☐ Disagree 48-4 ☐ Strongly disagree 48-5

Appendix B

Revised Appraisal System Design Preference Poll Questionnaire

INSTRUCTIONS: Read the descriptions of the Administrative only, Developmental only, Combined, and Separated appraisal systems and then rate each on the appraisal system preference scales (pages 2 and 3).

Administrative Appraisal Only

In an administrative appraisal system, the supervisor compares information across ratees. The supervisor seeks information about how others have performed similar tasks. The method of comparison is based on relative standing or ranking. When a supervisor has only one or a few ratees to rate, ratings of several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The supervisor's evaluation results in a single rating of the ratee's performance. This single rating is used as a basis for making administrative decisions such as merit increases and promotion. The appraisal discussion looks back at how the ratee has performed over the past rating period.

Developmental Appraisal Only

In a developmental appraisal system, the supervisor compares information about ratee performance across different work dimensions. The supervisor seeks information about how the ratee performed different tasks relative to work standards. The supervisor's evaluation results in ratings along different work dimensions. These dimensional ratings are used to identify areas of performance strength and weakness. This information is then used to prepare developmental action plans to improve the ratee's future performance in specific dimensions.

Combined Appraisal

(Both administrative and developmental appraisals conducted at the *same time*):

A combined appraisal system combines the administrative and developmental appraisals, with both conducted together. The supervisor looks at performance across ratees and across tasks to produce a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. Both types of ratings are discussed in the same session.

Separated Appraisal

(parallel administrative and developmental appraisals conducted at *different times*):

In a separated appraisal system, the administrative and developmental appraisals are conducted a few months apart so as to minimize any conflict in the supervisor roles of judge and coach. The supervisor looks at performance across ratees and across tasks to produce both a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. These two different types of ratings are discussed in separate sessions a few months apart.

Combined Appraisal

(Both administrative and developmental appraisals conducted at the *same time*):

A combined appraisal system combines the administrative and developmental appraisals, with both conducted together. The supervisor looks at performance across ratees and across tasks to produce a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. Both types of ratings are discussed in the same session.

1. The combined system would be a source of great satisfaction.
   ☐ Strongly agree 11-1 ☐ Agree 11-2 ☐ Neutral 11-3 ☐ Disagree 11-4 ☐ Strongly disagree 11-5
2. The combined system would produce fair evaluations.
   ☐ Strongly agree 12-1 ☐ Agree 12-2 ☐ Neutral 12-3 ☐ Disagree 12-4 ☐ Strongly disagree 12-5
3. The combined system would produce accurate evaluations.
   ☐ Strongly agree 13-1 ☐ Agree 13-2 ☐ Neutral 13-3 ☐ Disagree 13-4 ☐ Strongly disagree 13-5
4. The combined system would be a good way to assess staff performance.
   ☐ Strongly agree 14-1 ☐ Agree 14-2 ☐ Neutral 14-3 ☐ Disagree 14-4 ☐ Strongly disagree 14-5
5. The combined appraisal system would be my first choice.
   ☐ Strongly agree 15-1 ☐ Agree 15-2 ☐ Neutral 15-3 ☐ Disagree 15-4 ☐ Strongly disagree 15-5

Separated Appraisal

(parallel administrative and developmental appraisals conducted at *different times*):

In a separated appraisal system, the administrative and developmental appraisals are conducted a few months apart so as to minimize any conflict in the supervisor roles of judge and coach. The supervisor looks at performance across ratees and across tasks to produce both a single rating and dimensional ratings. When a supervisor has only one or a few ratees to rate, single ratings from several supervisors are pooled and reviewed at the next level of supervision to determine relative standing or ranking. The single rating looks at past performance over the rating period and is used to make administrative decisions. The dimensional ratings are used to identify areas of performance strength and weakness and to discuss improvement in specific work dimensions. These two different types of ratings are discussed in separate sessions a few months apart.

1. The separated system would be a source of great satisfaction.
   ☐ Strongly agree 16-1 ☐ Agree 16-2 ☐ Neutral 16-3 ☐ Disagree 16-4 ☐ Strongly disagree 16-5
2. The separated system would produce fair evaluations.
   ☐ Strongly agree 17-1 ☐ Agree 17-2 ☐ Neutral 17-3 ☐ Disagree 17-4 ☐ Strongly disagree 17-5
3. The separated system would produce accurate evaluations.
   ☐ Strongly agree 18-1 ☐ Agree 18-2 ☐ Neutral 18-3 ☐ Disagree 18-4 ☐ Strongly disagree 18-5
4. The separated system would be a good way to assess staff performance.
   ☐ Strongly agree 19-1 ☐ Agree 19-2 ☐ Neutral 19-3 ☐ Disagree 19-4 ☐ Strongly disagree 19-5
5. The separated appraisal system would be my first choice.
   ☐ Strongly agree 20-1 ☐ Agree 20-2 ☐ Neutral 20-3 ☐ Disagree 20-4 ☐ Strongly disagree 20-5

Ranking of Alternatives

6. How do you rank your preference for the four alternative performance appraisal systems along with the alternative of adopting none of these systems? Put a "1" by your first choice, a "2" by your second choice, a "3" by your third choice, a "4" by your fourth choice, and a "5" by your last choice. *Do not include any tie ranks.*

	Rank	Alternative
21	____	Administrative appraisal only
22	____	Developmental appraisal only
23	____	Combined (both administrative and developmental appraisals conducted at the same time)
24	____	Separated (parallel administrative and developmental appraisals conducted at different times)
25	____	None of these systems

7. Given a choice of either an Administrative Appraisal Only System or a Developmental Appraisal Only System, which would you choose? (*Check only one answer*)

□ Administrative Appraisal Only 26-1 □ Developmental Appraisal Only 26-2

8. Given a choice of either an Administrative Appraisal Only System or a Combined System (both administrative and developmental conducted at the same time) which would you choose? (*Check only one answer*)

□ Administrative Appraisal Only 27-1 □ Combined 27-2

9. Given a choice of either a Developmental Appraisal Only System or a Combined System (both administrative and developmental conducted at the same time) which would you choose? (*Check only one answer*)

□ Developmental Appraisal Only 28-1 □ Combined 28-2

10. Given a choice of either an Administrative Appraisal Only System or a Separated System (parallel administrative and developmental appraisals conducted at different times) which would you choose? (*Check only one answer*)

□ Administrative Appraisal Only 29-1 □ Separated 29-2

11. Given a choice of either a Developmental Appraisal Only System or a Separated System (parallel administrative and developmental appraisals conducted at different times) which would you choose? (*Check only one answer*)

□ Developmental Appraisal Only 30-1 □ Separated 30-2

12. Given a choice of either a Combined System (both administrative and developmental conducted at the same time) or a Separated System (parallel administrative and developmental appraisals conducted at different times) which would you choose? (*Check only one answer*)

□ Combined 31-1 □ Separated 31-2

Information About You

13. Which of the following best describes your functional classification?

□ Academic administrator 32-1 □ Professional staff 32-2 □ Biweekly staff 32-3

14. Which of the following best describes what your role would be if an appraisal system was adopted?

□ Rater only 33-1 □ Both rater and ratee 33-2 □ Ratee only 33-3

15. Which of the following best describes your age in whole years (31 and 9 months = 31)?

□ Below 32 34-1 □ 32 to 39 34-2 □ 40 to 47 34-3 □ 48 to 52 34-4 □ 53 and above 34-5

16. Which of the following best describes your length of service at IUSB in whole years (1 year and 9 months = 1 year completed)?

□ 1 year or less 35-1 □ 2 to 3 years 35-2 □ 4 to 6 years 35-3 □ 7 to 10 years 35-4 □ 11 or more years 35-5

17. If an appraisal system was adopted which included a single rating (e.g., to relate pay to performance), what would you expect your rating to be? *(Skip this question if you are an academic administrator)*

36-1 □ Unsuccessful (fails to meet a significant number of performance standards)
36-2 □ Needs improvement (falls below a few important performance standards)
36-3 □ Successful (meets all performance standards)
36-4 □ Exceeds expectations (surpasses most performance standards)
36-5 □ Outstanding (surpasses all performance standards by a wide margin)

Appendix C

Cover Letter

INDIANA UNIVERSITY SOUTH BEND

TO: Professional staff, Biweekly Clerical/Technical staff, and Academic Administrators who supervise staff

FROM: John Hundley, Director of Human Resources

DATE: February 5, 2001

SUBJECT: Performance Appraisal System Preference Poll

Currently IUSB bases annual increases for staff on an across-the-board percentage plus range penetration adjustments. (Range penetration adjustments are based on length of service up to five years.) Because IUSB is exploring the possibility of basing annual increases at least partially on merit, a Performance Evaluation and Merit Pay Task Force was formed. The Task Force conducted two surveys and used the responses to develop specifications for a comprehensive performance appraisal and merit pay system. The proposed specifications were presented to the Professional Staff Council, and Biweekly Staff Council, the Academic Administrative Council, the Chancellor's Staff, and to several divisional meetings for discussion. Some meeting participants suggested eliminating the developmental aspects of the proposed system while others suggested eliminating the administrative aspects. There was no consensus to adopt the proposed system specifications.

To resolve this dilemma, I am conducting a performance appraisal system preference poll which asks prospective users to express their preferences for several appraisal system alternatives. The preference poll is intended to help select a appraisal system which best fits the needs of those who would be affected by a such system. Chancellor Perrin has approved the conducting of this preference poll. I am also using this research effort for a Master's thesis to fulfill the requirements for a Master of Arts in Applied Psychology. Associate Professor of Psychology, Frank Fujita is chair of the thesis committee. This study has been reviewed for compliance with regulations of the University's Committee for the Protection of Human Subjects.

You should be able to complete the preference poll in 20 minutes or less. Your responses are confidential: the questionnaire does not ask for information that could be used to identify you as an individual. Please complete the questionnaire and return it to me in Human Resources (Administration Building, 244C) via campus mail by March 2, 2001. The summary of the preference poll findings will be available to interested individuals.

Please take the time to complete the attached preference poll. Not only will you help me with my research, your input will help determine IUSB's merit pay direction. The response rates to the first two surveys were quite low: 34.5% and 28.4%. These low response rates may be one of the reasons that the proposed system specifications did not reflect a consensus. Again, please respond as the findings will influence the systems used to evaluate staff performance and to determine staff merit pay.

VITA
JOHN R. HUNDLEY

CONTACT INFORMATION

ADDRESS: 52262 Chatem Court
Granger, IN 46530
OFFICE PHONE: (219) 237-4398
HOME PHONE: (219) 273-8524
E-MAIL: JHundley@IUSB.EDU

EXPERIENCE

INDIANA UNIVERSITY SOUTH BEND, *South Bend, IN* **1992-Present**
Director of Human Resources and Adjunct Instructor In Management
Indiana University South Bend is one of 8 campuses in the IU system with 7.500 students and a staff of roughly 500. Human Resources has a staff of 3 and reports to the Chancellor.

- Installed new staff compensation system despite predecessors' history of 3 prior failed attempts.
- Conducted multiple regression analysis and facilitated faculty pay equity task force.
- Negotiated annual agreements with AFSCME Local 1477 and resolved all South Bend grievances without arbitration.
- Received award for labor-management cooperation from local labor-management organization.
- Reduced employment advertising costs 40% with minimal impact on responses by using fewer words and fewer runs.
- Perfected employment screening questionnaire technique and coauthored two articles, one published twice.
- Initiated series of well received retirement and financial planning seminars.
- Developed highly praised benefits summary.
- Successfully defended 7 employment discrimination charges.
- Taught senior Personnel Research and Measurement and graduate Human Resource Management courses.
- Participated in planning "Conversations on Race", cited as a promising practice on the White House web site

GENERAL AMERICAN LIFE INSURANCE COMPANY, *St. Louis, MO* **1965-1990**
Human Resources Vice President
General American was a Fortune top 50 life insurer marketing individual and group life and health insurance as well as reinsurance, pensions, and annuities. The firm also administered Medicare Part B for the federal government. The firm employed 2,150 in St. Louis, 350 in field offices and 500 in subsidiary companies. Human Resources had a staff of 26 and for 10 years also included the mail & supply department with 36 employees.

- Operated Human Resources with staff 30% below industry average.
- Selected and developed 13 managers; 5 promoted into management outside HR.
- Coordinated the selection of a strategic planning consultant for top management resulting in restructuring by business units.
- Installed the Hay job evaluation system, evaluating 1,000 jobs by target date.
- Guided major organization changes for senior management by advising on job design, writing descriptions and finalizing pay.
- Conceived and installed group benefits hot line with 50+ recorded messages responding to employee concerns 24 hours a day.
- Reduced company's post retirement medical liability by recommending and implementing service eligibility requirements.
- Created a PC based, interactive retirement benefits estimator used by employees.
- Successfully defended 60+ employment discrimination charges over 25 years.
- Developed 20 affirmative action programs accepted by the OFCCP.
- Recommended and implemented management succession program which achieved senior management's needs.
- Developed training and development curriculum for each layer of management.
- Managed employment offices which filled non-exempt jobs in under 14 days (50% below industry average).
- Measured hiring quality by interviewer resulting in 95+% of new hires meeting or exceeding supervisors' expectations.
- Installed mainframe HR system which achieved earlier than projected payback.
- Wrote over 50 programs to generate reports and labels saving 400 staff hours.

WASHINGTON UNIVERSITY, University College, *St. Louis, MO* **1966-1978**
Adjunct Instructor In Management

- Taught courses in Personnel Management and Introduction to Management.
- Wrote 5 case studies published in several textbooks.
- Consistently awarded excellent student evaluations.

EDUCATION

MA Applied Psychology, Indiana University South Bend (2001)

- Completed 48/42 hours; 4.0/4.0 GPA
- Received Masters in Applied Psychology Award for Excellence
- Named to Psi Chi (psychology honorary)
- Thesis: User preferences for appraisal systems with different purposes

MBA, Washington University (1965)

- Named to Beta Gamma Sigma (business honorary)
- Awarded half tuition fellowship

BSBA, Washington University (1964)

- Graduated with honors

Development Dimensions, Inc. (1995) - trained assessor for three different client assessment centers (conducted as self-employed consultant)

LOMA Executive Development Program (1985)

PROFESSIONAL ASSOCIATIONS

College and University Personnel Association
Committee for Labor and Management Relations (South Bend)
American Psychological Association (graduate student member)
Society for Industrial and Organizational Psychology (graduate student member)
Chicago Industrial-Organizational Psychologists (graduate student member)
Human Resources Management Association of Greater St. Louis (Past President)
Society for Human Resources Management (Past District Director)

CIVIC ACTIVITIES

St. Joseph County Chapter of the American Red Cross Board (past Board Secretary and Chair, Volunteer & Staff Development Committee)
Confluence St. Louis (past Chair, Public Education & Economic Development Committee)
Leadership St. Louis Participant
Missouri Goodwill Industries Board (past Member)
St. Louis Chapter of the American Red Cross Board (past Chair, HR Committee)
Visiting Nurse Association of Greater St. Louis Board (past Secretary, Chair, HR Committee)
United Way of Greater St. Louis (past Chair, Personnel Committee)
Private Industry Council of St. Louis/St. Louis County (past Member)
Elected to political office (1973-1980)

PUBLICATIONS

"Path Analysis of a University's Professional Staff Pay", Proceedings of the Midwest Academy of Management (Chicago) April, 2000

"Developing a Selection System","The Promotion of Melba Moore" Cases, Incidents, and Experiential Exercises in Human Resources Management (Third Edition), Raymond L Hilgert, Cyril C. Ling, and Edwin Leonard (Dame, a Division of Thompson Learning, 2000)

"Conflict And Collaboration In The Hiring Process", Proceedings of the Midwest Management Society, (Chicago) March, 1997, coauthored with Steven D. Norton

"Written Case Evaluation Form", Instructor's Manual: Cases and Experiential Exercises in Human Resources Management (Second Edition), Raymond L Hilgert and Cyril C. Ling (Prentice Hall, 1996)

"Equal Employment Opportunity and ADA Implications of Screening and Selection", Proceedings of the Midwest Academy of Management, (St. Louis) April, 1995, and the CUPA Journal (Summer, 1995 Vol. 46, No. 2) coauthored with Steven D. Norton

"The Use of Questionnaires For Screening Job Applicants", Proceedings of the Midwest Management Society, (Chicago) March, 1995, coauthored with Steven D. Norton

"Age Discrimination Cannot Continue", HR Magazine, October, 1992

"The Training Manager's Objectives", "Global United (B): The Older Supervisor", "Hazard Protection Insurance Company", Cases And Exercises In Human Resources Management, Raymond L Hilgert, Sterling H. Schoen, and Cyril C. Ling, (Allyn and Bacon, 1990)

"General American Evaluates Change, Readies For More", LOMA Resource, July-August, 1986 edited by Laura Godfrey.

"Willis B. Johnson", Cases and Policies in Personnel/Human Resources Management (4th Edition), Raymond L. Hilgert, Sterling H. Schoen, and Joseph W. Towle, (Houghton Mifflin Company, 1982)

"Planning Personnel's Contribution to Productivity", LOMA Resource, March-April, 1980

"Developing Quality Supervision", LOMA Resource, January-February, 1977

"Listening Posts", LOMA Resource, May-June, 1976 and Personnel, July-August, 1976

"Supervision: the Weak Link in Flexible Work Scheduling?", The Personnel Administrator, January, 1975 coauthored with Raymond L. Hilgert.